Genome sizes and repeatome evolution in zoantharians (Cnidaria: Hexacorallia: Zoantharia)

Chloé Julie Loïs Fourreau [Corresp., 1], **Hiroki Kise** [1, 2], **Mylena Daiana Santander** [3], **Stacy Pirro** [4], **Maximiliano M Maronna** [5], **Angelo Poliseno** [1], **Maria EA Santos** [1, 6], **James Davis Reimer** [1, 7]

[1] Molecular Invertebrate Systematics and Ecology (MISE) Lab, Graduate School of Engineering and Science,, University of the Ryukyus, Nishihara, Okinawa, Japan

[2] AIST Tsukuba Central, Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan

[3] Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

[4] Iridian Genomes, Bethesda, United States

[5] Faculdade de Ciências, Universidade Estadual Paulista (UNESP), Bauru, Brazil

[6] Okinawa Institute of Science and Technology, Onna, Okinawa, Japan

[7] Tropical Biosphere Research Center, University of the Ryukyus, Nishihara, Okinawa, United States

Corresponding Author: Chloé Julie Loïs Fourreau
Email address: chloisf@gmail.com

Across eukaryotes, large variations of genome sizes have been observed even between closely related species. Transposable elements as part of the repeated DNA have been proposed and confirmed as one of the most important contributors to genome size variation. However, the evolutionary implications of genome size variation and transposable element dynamics are not well understood. Together with phenotypic traits, they are commonly referred to as the “C-value enigma”. The order Zoantharia are benthic cnidarians found from intertidal zones to the deep sea, and some species are particularly abundant in coral reefs. Despite their high ecological relevance, zoantharians have yet to be largely studied from the genomic point of view. This study aims at investigating the role of the repeatome (total content of repeated elements) in genome size variations across the order Zoantharia. To this end, whole-genomes of 32 zoantharian species representing five families were sequenced. Genome sizes were estimated and the abundances of different repeat classes were assessed. In addition, the repeat overlap between species was assessed by a sequence clustering method. The genome sizes in the dataset varied up to 2.4X fold magnitude. High correlations between genome size, repeated DNA content (Pearson’s R=0.73, p=0.00052), and transposable elements, respectively, were found, suggesting their involvement in the dynamics of genome expansion and reduction. In all species, Long Interspersed Nuclear Elements and DNA transposons were the most abundant identified elements. These transposable elements also appeared to have had a recent expansion event. This was in contrast to the comparative clustering analysis which revealed species-specific patterns of satellite elements’ amplification. In summary, the

genome sizes of zoantharians likely result from the complex dynamics of repeated elements. Finally, the majority of repeated elements (up to 80%) could not be attributed to a known repeat class, highlighting the need to further investigate non-model cnidarian genomes. More research is needed to understand how repeated DNA dynamics relate to zoantharian evolution and their biology.

Genome sizes and repeatome evolution in zoantharians (Cnidaria: Hexacorallia: Zoantharia)

Chloé Julie Loïs Fourreau[1*], Hiroki Kise[1,2], Mylena Daiana Santander[3], Stacy Pirro[4], Maximiliano Manuel Maronna[5], Angelo Poliseno[1], Maria Eduarda Alves Santos[1,6] and James Davis Reimer[1,7]

[1]Molecular Invertebrate Systematics and Ecology (MISE) Lab, Graduate School of Engineering and Science, University of the Ryukyus, Nishihara, Okinawa, Japan

[2]Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central, Tsukuba, Ibaraki, Japan

[3]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil

[4]Iridian Genomes, Bethesda, Maryland, United States

[5]Faculdade de Ciências, Universidade Estadual Paulista (UNESP), Bauru, SP, Brazil

[6]Okinawa Institute of Science and Technology, Okinawa, Japan

[7]Tropical Biosphere Research Center, University of the Ryukyus, Okinawa, Japan

*Corresponding author:
Chloé Julie Loïs Fourreau
University of the Ryukyus, Senbaru 1, Nishihara, Okinawa 903-0129, Japan
chloisf@gmail.com

Abstract

Across eukaryotes, large variations of genome sizes have been observed even between closely related species. Transposable elements as part of the repeated DNA have been proposed and confirmed as one of the most important contributors to genome size variation. However, the evolutionary implications of genome size variation and transposable element dynamics are not well understood. Together with phenotypic traits, they are commonly referred to as the “C-value enigma”. The order Zoantharia are benthic cnidarians found from intertidal zones to the deep sea, and some species are particularly abundant in coral reefs. Despite their high ecological relevance, zoantharians have yet to be largely studied from the genomic point of view. This study aims at investigating the role of the repeatome (total content of repeated elements) in genome size variations across the order Zoantharia. To this end, whole-genomes of 32 zoantharian species representing five families were sequenced. Genome sizes were estimated and the abundances of different repeat classes were assessed. In addition, the repeat overlap between species was assessed by a sequence clustering method. The genome sizes in the dataset varied up to 2.4X fold magnitude. High correlations between genome size, repeated DNA content (Pearson’s R=0.73, p=0.00052), and transposable elements, respectively, were found, suggesting their involvement in the dynamics of genome expansion and reduction. In all species, Long Interspersed Nuclear Elements and DNA transposons were the most abundant identified elements. These transposable elements also appeared to have had a recent expansion event. This was in contrast to the comparative clustering analysis which revealed species-specific patterns of satellite elements’ amplification. In summary, the genome sizes of zoantharians likely result from the complex dynamics of repeated elements. Finally, the majority of repeated elements (up to 80%) could not be attributed to a known repeat class, highlighting the need to further investigate non-model cnidarian genomes. More research is needed to understand how repeated DNA dynamics relate to zoantharian evolution and their biology.

Keywords: C-value enigma; Genome size; Repeated DNA; Transposable elements; Zoantharia

Introduction

In biology, the ideas that some organisms are more complex than others and that evolution is directed towards progress and increasing complexity have oriented many research topics. In fact, these lines of thinking form the premise of a question that has long puzzled genome scientists, the “C-value paradox”. This paradox is related to the lack of correlation between the C-value, i.e. the size of a species’ haploid genome, and the expected complexity of an organism (Elliot & Gregory, 2015). Although the current understanding of evolution does not support the view of complexification in certain taxa, discrepancies described by the C-value paradox underline the confusing variations of genome sizes. Indeed, despite genome sizes being in most cases remarkably constant within species (Swift, 1950), intraspecific variation is well recognized and large variations exist between closely related species. The discrepancies between genome size, phenotype complexity and genomic content was reframed by the discovery of large amounts of repetitive DNA in genomes (Gregory, 2005). However, this finding raised even more questions regarding the impact of these repetitive elements (including both protein-coding and non-coding sequences) on evolutionary dynamics: mechanisms (e.g., amplification), historical processes (gain or loss of DNA content), and how repeats may relate to organismal and ecological traits. The set of questions that have risen from deciphering the “C-value paradox” are now collectively referred to as the “C-value enigma” (Gregory, 2005).

The development of next-generation sequencing along with tools dedicated to the annotation of specific repeated elements has allowed to describe and identify in detail various classes of genome repetitive elements. Currently, they are classified into two large groups based on their potential for mobility; tandem repeats and transposable elements. Tandem repeats include satellites, microsatellites, and rDNA (Bourque *et al.,* 2018). On the other hand, transposable elements (TEs) are capable of moving within a genome and can be distinguished into two classes based on their transposition mechanisms (Wicker *et al.,* 2007). Class I TEs, also known as retrotransposons, insert themselves by reverse transcription; they include LTRs (Long Terminal Repeats), LINEs (Long Interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements). Transposable elements of class II encode for a transposase, an enzyme that performs transposition. These elements include Helitrons, Maverick and other DNA transposons subcategories (Wicker *et al.,* 2007).

Repeated elements have been referred to as “junk DNA” and were initially thought to be neutral with regards to genome evolution. However, their dynamics can have large implications on the genome and species biology. For example, TEs can have adverse effects on their host by causing cancer (Bourque *et al.,* 2018), including transmissible cancers through horizontal transfers in the marine environment (Metzger *et al.,* 2018). Furthermore, TEs can lead to sequence polymorphism and gene diversification through genomic rearrangements and mediation of gene expression. As examples of this, transposable elements have promoted the diversification of opsins in the amphioxus genome (Pantzartzi *et al.* 2018), and a TE insertion event gave rise to the dark

morphotype of the peppered moth (Van't Hof *et al.,* 2016). Finally, TEs have been associated with hybrid defects, and are thus potentially involved in the speciation process (Serrato-Capuchina & Matute, 2018). For all these reasons, repeated elements are relevant to the understanding of species biology and evolution.

To better understand repeated elements, genome sizes, and their implications in organism evolution, further research on understudied groups is necessary (Elliot & Gregory, 2015). Hotaling *et al.* (2021) highlighted important taxonomic biases in genome sequencing projects, showing large research bias in favor of vertebrates. This is also true for the study of repeatome and genome sizes, as many groups still lack basic genome size information. In phylum Cnidaria, the first study documenting genome sizes across a wide taxonomic scope was published in 2017 by Adachi *et al*. While most cnidarians seem to have relatively small genomes (e.g., mean C values: 0.70 pg for Anthozoa, 0.46 pg for Scyphozoa, and 1.20 pg for Hydrozoa) compared to other metazoans, there is a >13-fold variation in their genome diversity, (from 0.26 pg in scyphozoan *Sanderia malayensis* to 3.56 pg in hydrozoan *Agalma elegans*; Adachi *et al.,* 2017). However, more research is needed to fully understand the scope and diversity of genome size variation in Cnidaria. Zoantharians represent one of the several taxa within the phylum for which no estimates of genome sizes have yet been published.

The order Zoantharia Rafinesque, 1815 is considered the earliest branching hexacorallian group (Quattrini *et al.,* 2020) and their study harbors important implications for the evolution of cnidarian traits including skeleton production (Quattrini *et al.,* 2020), symbioses, coloniality, and development (Hirose et al., 2011). Zoantharians are extensively distributed in subtropical and tropical oceanic regions and inhabit intertidal zones to the deep sea (Santos *et al.* 2019) and, in certain environments, can be dominant (Yang *et al*., 2013). In suborder Brachycnemina, most species establish symbiosis with photosynthetic dinoflagellates of the family Symbiodiniaceae, and azooxanthellate species (i.e., that do not host Symbiodiniaceae) are thought to have lost this relationship (Irei *et al.,* 2015). On the other hand, zoantharians of the suborder Macrocnemina are usually azooxanthellate, and epizoic on a range of marine invertebrates, including sponges, hermit crabs, molluscs, annelids, urchins, and several different groups of anthozoans (Kise *et al.,* 2019). In addition, some species of zoantharians are known to produce palytoxin, one of the most potent toxic compounds known from the marine environment (Aratake *et al.* 2016), and present potential therapeutical applications. The phylogenetic relationships of zoantharians are currently debated and have been the focus of a few phylogenomic reconstructions; examples include a detailed phylogeny of genus *Palythoa* from eZRAD (Dudoit *et al.,* 2021), the placement of Zoantharia within Cnidaria from ultra-conserved elements (Quattrini *et al.,* 2020), and the phylogeny of Zoantharia from mitochondrial genome datasets (Poliseno *et al.,* 2020). Some of these phylogenies (Poliseno *et al*., 2020*;* Quattrini *et al.,* 2020)

together with previous single marker phylogenetic results indicate that the taxonomy of zoantharians should be revised, since Brachycnemina is nested within Macrocnemina (Sinniger *et al.,* 2005).

Despite the high relevance of zoantharians in terms of evolution, ecology and biochemical potential, this group has yet to be well studied from a general genomic point of view. To fill this gap we investigated the genomes of 32 species of zoantharians, spanning 11 genera of the order and 5 out of 9 families. We present newly sequenced data for 17 of those species. From this recent and mostly unexplored molecular resource, we aimed to (1) expand present mitochondrial data via increased taxon sampling to test the current view of zoantharian phylogeny, (2) provide baseline data on zoantharian genomes with regards to genome sizes and repeatomes, and (3) assess the relative importance of different repeated DNA classes in genome size evolution in the order.

Material and Methods

Sampling and sequencing

Thirty-two specimens of zoantharians were gathered from SCUBA diving, scientific deep-sea expeditions, and museum collections between 1982 and 2019, from the Pacific Ocean, the Caribbean Sea and the South African coast of the Indian Ocean (Table 1). These specimens were fixed in 99% ethanol and kept at -20°C before 30 of them were sent to Iridian Genomes (Bethesda, USA) for whole-genome sequencing. DNA was extracted using the Qiagen DNeasy kit following manual's instructions. The sequencing platform, Illumina Hi-Seq, generated approximately 60 million paired-end reads of a size of 150 bp per specimen. Genome data for 11 brachycnemic zoantharian specimens (Santos *et al.*, 2023) and the 5 *Epizoanthus* species in the scope of the present paper have been already presented (Kise et al., 2023a; Santos *et al.*, 2023). In the case of the sample of *Palythoa mizigama*, DNA was extracted by CTAB-based protocol and sequenced at the NovoGene Hong Kong facility using the Illumina HiSeq X Platform (NEBNext® DNA Library Prep Kit was used for library construction (350 pb insert size, 150 pb read length), including size selection and PCR-enrichment, with a total input amount of 1.0 µg DNA). For the whole genome sequencing of *P. tuberculosa* ~1 µg of genomic DNA was sent to Admera Health (South Plainfield, NY). Genomic library was prepared using a Kapa® HyperPrep kit (Roche) and it was sequenced on Illumina Hi-Seq platform using a 150 pair-end chemistry.

The sequencing experimental data are available on the Sequence Read Archive with accession numbers as reported in Table 1. All SRA paired-end reads were downloaded onto the National Institute of Genetics Supercomputer Cluster (https://sc.ddbj.nig.ac.jp/en) to proceed with subsequent bioinformatic analyses. Before any analyses, the samples were quality-checked using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Mitochondrial genome assembly and phylogeny

Mitochondrial genome assembly

Before the mitochondrial assembly, paired-end reads adapter sequences were removed in Trimmomatic v. 0.39 with default parameters (Bolger *et al.,* 2014). Then, mitochondrial genomes (mtDNA) were assembled *de-novo* with NOVOPlasty v. 3.8.3 (Dierckxsens *et al.,* 2017), with a k-mer size comprised from 29 to 33. A partial COI sequence (~780bp) from *Palythoa tuberculosa* (GenBank accession number: MH013403) was chosen as seed for the assembly of the majority of the samples, yet for others we used the whole sequence of phylogenetically close mt-genomes retrieved from GenBank or the sequence of some protein-coding genes such as for instance COI and COIII. Although the assembly was performed *de novo*, the input of a reference genome facilitates the process, and therefore, the mitogenome of *Palythoa heliodiscus* was used (Chi & Johansen, 2017; NC035579). To identify the gene composition and order, mitochondrial genomes were circularized and annotated in Geneious v.8.1.9. (Kearse *et al.* 2012). This was done using the Predict and Annotate tool by comparing mitogenomes with a reference mitogenome annotation of *Palythoa heliodiscus* (MN863593) and other zoantharian mt-genomes from Poliseno *et al.* (2020). Protein-coding sequences with >75% similarity to a gene in the reference were assigned to the corresponding gene.

Mitochondrial genome phylogeny

To infer the evolutionary relationships of zoantharians, phylogenetic trees were inferred based on mitochondrial protein coding genes. Thirteen genes (COI, COII, COIII, CYTB, ATP6, ATP8, NAD1, NAD2, NAD3, NAD4, NAD4L, NAD5, NAD6) were retrieved from each genome and aligned individually with MUSCLE (Edgar, 2004). Additional mitogenomes available from the literature and incorporated in the dataset are listed in Table 2. The antipatharians *Stichopathes luetkeni* (Kayal *et al.,* 2013) and *Myriopathes japonica* (Kwak, Choi *et* Hwang, unpublished) mitogenome assemblies were used as outgroups in the phylogenetic trees. The thirteen alignments were concatenated in Sequence Matrix v.1.8 (Vaidya *et al.,* 2011), resulting in 11,933 bp matrix. The best fitting evolutionary model of each gene was assessed with MEGA X (Kumar *et al.,* 2016) using the AIC criterion (Akaike, 1973).

Based on the concatenated alignment, phylogenetic trees were computed following the maximum-likelihood method in RAxML-NG using the command `--all` (Kozlov *et al.,* 2019), which comprises of an initial tree search step and a non-parametric bootstrapping step with node support estimated by 1000 replicates. Furthermore, a Bayesian phylogenetic tree was inferred with MrBayes v.3.2.7 (Ronquist & Huelsenbeck, 2003). Each Monte Carlo Markov Chain (MCMC) was sampled every 1000 steps during $10 \cdot 10^6$ generation cycles, and the first 25% of the trees were discarded as burn-in. Tree node parametric support was evaluated with the Bayesian posterior probabilities calculated during the analysis. For both, the maximum-likelihood and the Bayesian tree computations, partitions were set with the corresponding sequence evolution model of each gene.

Comparative genomic analyses

Genome sizes

To estimate genome sizes, the k-mer frequencies of previously trimmed reads were counted in Jellyfish (Marçais *et al.,* 2011) with the command `jellyfish-count` and the default k-mer size of 21. With the command `jellyfish-histo`, histograms were computed, then input in GenomeScope (http://qb.cshl.edu/genomescope), which estimates genome size based on the distribution of a given k-mer size.

Abundance and Annotation of repeat classes

The pipeline dnaPipeTE v.1.3.1 (Goubert *et al.,* 2015) was employed to assemble, annotate and estimate the abundance of repeated elements in each zoantharian genomic dataset. This software uses low coverage read samples to assemble representative contigs of repeats with Trinity v.2.5.1 (Grabherr *et al.,* 2013) and then, annotates the resulting contigs with Repeatmasker (Smit et al., RepeatMasker Open 4.0.7, http://www.repeatmasker.org) and RepBase (Bao *et al.,* 2015). The dnaPipeTE pipeline also estimates repeat abundances and the divergence of repeat copies to the assembled contigs via blastn (Altschul *et al.,* 1990). Both pieces of information are then used to estimate the landscape distribution of repeated elements, as a proxy of their relative age.

To ensure the sampling of repeated elements, reads were trimmed and removed with stricter parameters than the default Trimmomatic command. The chosen parameters demanded a minimum read length of 140 bp instead of the default 36 bp (MINLEN:120), as well as an average quality (SLIDNGWINDOW:4:20) below 20, instead of the default 15 (SLIDNGWINDOW:4:20).

To avoid misrepresenting the repeat composition, non-repeat sequences with high coverage must be filtered out of the dataset (Goubert *et al.*, 2015). The mitochondrial genomes previously assembled were removed from the trimmed reads using the script bbsplit.sh from bbmap package (Bushnell, 2014).

To produce comparable estimates of repeated elements between species, the “fixed read sampling size” method was used (as opposed to using genome coverage). To determinate the appropriate number of reads to sample, tests were runs by providing genome sizes, and with dnaPipeTE coverage options of 0.1, 0.2, 0.25, 0.3, 0.4, and 0.5 -fold for two of the datasets with largest genomes (*Palythoa tuberculosa* and *Umimayanthus chanpuru*) and one of the smallest (*Hydrozoanthus tunicans*). The resulting Trinity assemblies of annotated and unannotated contigs (annoted.fasta, unannoted.fasta, Trinity.fasta output files) were evaluated with the L50 metric using the bbtools script stats.sh (Table S1). Based on this, the optimal sample size (number of reads) was assessed using the formula C=(N*L)/G with C the coverage, N the number of reads, L the read length (150 bp) and G the genome size. To determine N, C was set as 0.4 based on the results of dnaPipeTE test runs (Table S1). To ensure that all

datasets were sufficiently sampled, G was input as the smallest genome size recovered, from *Palythoa mizigama* (G=286,669,957 bp). Based on this calculation, the read sampling size was fixed to 764453 for all species. In addition, the minimum size of contig to be included was set to 400bp.

Finally, the output files “Counts.txt” and “reads_landscape” of dnaPipeTE analysis, containing counts of each annotated repeat class, per species, were employed for statistical analyses. These files were remanipulated to create Figures 2, 3 and 4 and the corresponding tables are available as supplementary materials (Supplementary Tables S2, S3, S4).

Repeat clustering and comparative composition between species

To analyze whether sequences of different classes of repeated DNA were shared between zoantharian species, a comparative analysis was performed with RepeatExplorer2 (Novák *et al.,* 2020). This pipeline allows the clustering, quantification and annotation of repeats from unassembled short reads, on the web interface Galaxy. It was employed in comparative mode for the 18 species with available genome size information. Pre-processing was performed on RepeatExplorer2 as described in Protocol 2 of the pipeline manual (Novák *et al.,* 2020), including the subsampling of 500,000 reads, the interlacing of paired-end reads and the concatenation of all species reads into a single file. The clustering of the reads was performed in comparative mode using the Repeat Explorer database in Metazoa version 3.0 and default parameters. In this process, RepeatExplorer2 performs the clustering of the reads regardless of the species they belong to. Therefore, similar reads of different species clustered together, representing groups of repeated elements that are shared between different species. On the other hand, clusters that were composed of reads from a single species were considered specific repeats. The RepeatExplorer2 clustering outputs a list of superclusters along with their annotation. Because of conflicts during the annotation process, each supercluster annotation was reviewed and manually corrected as advised by Novák *et al.* (2020). Clusters that could not be assigned to a repeat type were viewed in tablet (Milne *et al.,* 2013) and the contig with the most important number of reads was inspected. When the reads at the tip portions of the contig showed high polymorphism, the cluster was considered a mobile element (Novák *et al.* 2020), as this structure represents several different insertion sites of a transposon. Finally, in order to visualize clusters that were shared or not shared between species, the corrected version of the cluster annotation file `cluster_table.csv` was input in Repeat Explorer visualizing tool. This comparison was generated in a raw version as well as a version where cluster abundance was normalized by genome size.

Statistical analyses and visualization

To be able to relate evolutionary history with repeat abundance and genome sizes, a cladogram was drawn based on the topology of phylogenetic trees computed with the mitochondrial datasets, pruning the branches of specimens without genome size data in TreeViewer v.2.0.1 (https://treeviewer.org/). As the mitogenome of *Umimayanthus*

chanpuru could not be reconstructed, this species was placed on the cladogram with *Umimayanthus nakama*, based on phylogenetic reconstructions from the literature (Montenegro *et al.,* 2015). Results were visualized with the R package ggtree (Yu, 2020).

To evaluate whether genome sizes were correlated to total repeated elements or transposable elements, a regression analysis was performed with the `lm` function in R (R Core Team, 2021). To ensure that datasets meet the conditions required for the Pearson correlation test, the plots (Residuals vs Fitted, Scale-Location, Normal Q-Q and Normality vs Leverage plots) produced by the `lm` function were examined. Additionally, the normality of residuals distribution was assessed with a Shapiro-Wilk test. Final plots were generated with `ggscatter` from the ggplot2 package (Wickham, 2016). The correlation between genome size and each TE class was evaluated with Spearman's rank correlation test, and plots suggesting a linear relationship were further evaluated with a Pearson's test.

Results

Mitochondrial genomes and phylogeny

Of the 32 mitogenomes for which assembly was performed, 29 of them could be assembled into a single circularized contig. Two species, *Umimayanthus chanpuru* and *Epizoanthus planus*, failed to generate a successful assembly. The processing of *Paleozoanthus reticulatus* resulted in a partial assembly of seven contigs, of which only four genes could be retrieved, ATP6, ATP8 and ND4L on one contig (Genbank accession: OQ843460) and COX1 on another (OQ848443).
All other mitochondrial genomes were circularized and presented the complete gene set, displaying the same gene arrangement as described by Chi & Johansen (2017); COII, NAD4, NAD6, CYTB, COIII, COI (with an intron), NAD4L, ATP8, ATP6, NAD2 and NAD5 including NAD1 and NAD3 gene copies in its intron. Mitochondrial genomes sizes ranged between 19386bp for *Epizoanthus rinbou* and 23133bp for *Umimayanthus parasiticus.* A table summarizing the sizes of all complete mtgenomes is available in the supplementary material (Table S2).

Sequence evolution models were HKY+G+I (Hasegawa *et al.,* 1985) for ND5 and ND4L, GTR+G+I (Tavaré, 1986) for ND1, ND2, ND3, ND4, ND6, CYTB, COIII, COI, ATP6, and T92+G+I (Tamura, 1992) was the best fitting model for COII and ATP8. Because T92+G+I was not available in MrBayes nor raxml-ng, the second best fitting model was employed for these two genes, in both cases HKY+G+I.

The phylogenetic reconstructions performed with Bayesian inference and maximum-likelihood methods (Fig. 1) found the suborder Brachycnemina to be monophyletic with high support (Bayesian posterior probabilities=1, maximum-likelihood bootstrap=100%). Conversely, Macrocnemina was retrieved as paraphyletic, containing Brachycnemina as the macrocnemic genus *Hydrozoanthus* which was sister to Brachycnemina. Families Sphenopidae, including the genera *Palythoa* and *Sphenopus*, and Zoanthidae, comprising *Zoanthus* and *Neozoanthus*, were respectively found as monophyletic. The

azooxanthellate, non-colonial species *Sphenopus marsupialis* was retrieved as a sister species to another azooxanthellate Sphenopidae*, Palythoa mizigama.* Similarly, *Hydrozoanthus* included a member of another genus, *Paleozoanthus reticulatus*, which was sister to *Hydrozoanthus gracilis*, with high support obtained only with the Bayesian inference (pp=0.99; bootstrap=66%).

Genome sizes and repeated elements content

Genome sizes estimates were obtained for 18 species (Supplementary table S3). While estimates were obtained for *Epizoanthus planus* (38,964,917 bp) and *Paleozoanthus reticulatus* (28,412,256 bp), these were considered unreliable based on the spectrum generated by GenomeScope, which did not point to a clear k-mer peak. Genome size estimates could not be obtained from sequencing data of 12 additional species. Genome size of zoantharians species ranged between 286 and 678 million base pairs (Mbp). The genera *Zoanthus, Umimayanthus* and *Hydrozoanthus* overlapped in range with genome sizes between 370 Mbp and 590 Mbp, and maximum disparities within genus of 160 Mbp. Genus *Palythoa*, however, comprised the maximum disparities at the scale of the order with a 2.4x fold variation and the maximum and minimum genome sizes, belonging respectively to *P. tuberculosa* and *P. mizigama* (Fig. 2C).

A range overlap in genome sizes between species in different genera was also apparent in the abundance of repeat reads, which accounted for 40 Mbp in several species (Fig. 2, Fig S1). The read abundance for each repeated element class and species are reported in Table S4. Despite similar total repeat abundances, the proportions of repeat classes seemed to vary (Fig. 2, Fig. S1). Of all identified repeats, up to 30 Mbp (~70% of total repeated elements) could not be attributed to a known repeat class (Fig. 2). The abundance of unannotated repeats seemed to reach higher proportions in the comparatively smaller genomes of *P. mizigama, H. tunicans*, and *H. antumbrosus.* TEs were more abundant than other repeated elements. In particular, LINEs and DNA elements were consistently the most abundant classes among zoantharian species (Fig. 2, Fig. S1). LINEs elements were, in all species, especially represented by the LINE/L2 family and Penelope elements, which reached respectively up to 20,000 and 10,000 copies (Fig. 3, Table S5). LINE/RTE-BovB elements were particularly abundant in *Zoanthus* species, reaching about 15,000 copies in *Z. solanderi*, while being under 5,000 copies in other genera. Congeneric species of the genus *Bergia* appeared to have similar genome sizes of about 530 Mbp, and almost identical compositions of repeated elements. The same was true for *H. tunicans* and *H. antumbrosus,* which both had genome sizes of 370 Mbp. Conversely, species of *Umimayanthus* and *Zoanthus* showed a nearly identical composition of repeated elements despite having different genome sizes (Fig. 2). At a higher taxonomic level, there was no evident pattern of differences between species of the suborder Macrocnemina and Brachycnemina, except for the fact that macrocnemic zoantharians had a higher abundance of rRNA repeats. However, the clade including Brachycnemina and *Hydrozoanthus* appeared to have higher number of SINEs elements copies, while these were almost completely lacking from other macrocnemic zoantharians.

Sphenopus marsupialis had a large amount of DNA/Maverick copies compared to other zoantharians (Fig. 3, Table S6).

Most of the transposable element landscapes showed a unimodal distribution with a spike of read abundance corresponding to a divergence of 0 to 2.5% from dnaPipeTE contig (Fig. 4). Abundance of TE reads increased gradually in *Zoanthus, Umimayanthus chanpuru* and *Palythoa tuberculosa*, while in other macrocnemic taxa, and in *S. marsupialis* and *P. mizigama,* most of the reads showed a peak at low divergences. DNA and LTR elements appeared to have a higher number of low divergence copies than LINEs in *S. marsupialis.* A few species displayed a bimodal distribution with increased number of LINEs elements at a high percentage of divergence. The second spike was stronger in *H. antumbrosus* which displayed an increased abundance of LINEs elements at a divergence of about 13%, while *H. tunicans,* its sister species according to the mitochondrial phylogeny (Fig. 1), did not show any other spike, and had relatively fewer LINEs elements at this degree of divergence. *Zoanthus solanderi* also displayed a small bump related to the activity of LINEs elements at ~25% of divergence, and a similar bump was also present but much dampened in a close-related species, *Z. gigantus.* In most species, DNA elements were as abundant as LINE elements at divergences higher than 2.5%. Conversely, in *Umimayanthus, Palythoa* and *Sphenopus* DNA elements appeared instead to be more important at divergences higher than 2.5%. At low divergences, LTR elements appeared to have higher abundances, whereas SINE elements disappeared, being at their peak abundance (~0.12% of genome) at 10% divergence. Landscapes of the same 18 species including lower level of repeated DNA classifications are available in Fig. S2.

Repeated elements clustering and comparative analysis among zoantharians

The repeated elements clustering in RepeatExplorer2 resulted in the analysis of 4,929,668 reads, of which ~60% were assigned to 354 superclusters, and 354 clusters. Total number of reads detected in each repeat class are summarized in Table S7. Many clusters were represented by all zoantharian species, in particular clusters displayed in Fig. 5 and Fig. S3 between cluster 349 and cluster 102, which were annotated as several different repeated element categories (45S, Maverick, LINEs and mobile elements). Other well-represented clusters among the zoantharian dataset were instead composed of unclassified elements, displayed between clusters 105 and 155 (Fig. 5, Fig. S3), which were found in increased abundance in *Zoanthus*. However, in general, clusters that were present among all zoantharian species did not seem to be found in high proportions with respects to genome size (Fig. 5). Clusters retrieved in larger number were mostly species-specific or shared among closely related species of the same genus. In particular, several closely related species with almost identical genome sizes displayed very similar clusters in high abundance. This includes the two *Bergia* species with clusters 155 to 212, *Z. solanderi* and *Z. gigantus* (clusters 229 to 317), and the closely related *H. antumbrosus* and *H. tunicans*, with mostly satellites and LINEs

elements (clusters 5 to 56). These groups of clusters corresponded essentially to satellite elements in the species pairs mentioned above. However, abundant clusters of LINEs were also shared among the two *H. antumbrosus* and *H. tunicans* (clusters 251 and 190) and among all *Zoanthus* species (cluster 29). 5S RNA was shared and particularly abundant in *Z. solanderi* and *Z. gigantus*. Conversely, several satellite clusters were found in high abundance in a single species only, mostly species displaying the highest genome size of their group (*H. sils, Z. pulchellus*, *S. marsupialis*) (Fig. 2, Fig. 5). *Z. pulchellus* and *Z. sociatus* had the highest genome sizes in *Zoanthus* (580 and 553Mb respectively, Fig. 2) but had different clusters amplified; cluster 213 in *Z. sociatus* contained 35 million repeats while cluster 16 had 25 million repeats in *Z. pulchellus* (Fig. 5).

Correlation tests between genome size and repeated elements

Pearson's correlation test showed a high correlation between the 18 genome sizes and the proportions of repeated elements, supported by an R of 0.73 and a highly significant p-value of 0.00052. A weaker but statistically significant correlation was found between genome sizes and the percentage of transposable elements (R=0.56, p=0.015), in which points appeared more dispersed (Fig. 6A, 6B). All Pearson correlation tests were made under the assumption that residuals followed a normal distribution, which was confirmed by the Shapiro-Wilk test, with p-values > 0.05.

On the other hand, no significant correlation was noted by the Spearman correlation tests between genome sizes and each separate repeat class. Satellite elements, simple repeats, SINEs, rRNA, Low complexity elements, Helitrons, LTRs, and other repeats had no pattern of variation related to genome size. However, the plots relating genome size, LINEs elements and unclassified repeats showed a slight slope, hinting at a linear relationship. Therefore, LINEs and unclassified repeats percentage were tested for a correlation with genome size using Pearson's correlation test, which showed statistically significant results (Fig. 6C, 6D).

Discussion

Mitochondrial genomes and phylogeny of order Zoantharia

This study extended the datasets of zoantharian mitochondrial genomes compared to previous works (Poliseno *et al.* (2020), adding thirty additional mitochondrial genomes from twenty-two species and including four genera that had not previously been reported. Mitochondrial gene rearrangements have been reported in the close-related subclass Ceriantharia (tube anemones; Stampar et al., 2019) and in all other orders of Hexacorallia, including Actiniaria (sea anemones; Johansen et al., 2021), Corallimorpharia (corallimorpharians; Lin *et al.,* 2014), and Scleractinia (stony corals; Lin *et al.,* 2014), but none were observed here for Zoantharia. Similar to zoantharians, a lack of variation in gene orders in black corals (order Antipatharia) has also been noticed. However, sampling of 18 species of the group lead to the discovery of mitogenomic rearrangements, in the form of a loss of COI intron in two families (Barrett *et al.,* 2020). The lack of evidence for gene rearrangements in zoantharians was also

hypothesized to be due to the reduced sampling effort (Poliseno *et al.,* 2020). Still, despite the increased taxon sampling of the present study, all mitochondrial genomes that could be completely assembled displayed the same gene order arrangement, which is identical to the one originally described by Sinniger et al., (2007) and Chi & Johansen (2017). As of this study, Zoantharia remains the only hexacoral order without gene rearrangements in the mitochondrial genome. Although sequencing more species in the future may uncover different mitochondrial gene arrangements, the current situation suggests that biological factors may constrain the structure of mitochondrial genomes in zoantharians, as has been previously suggested for antipatharians (Poliseno *et al.,* 2020). Our reconstructed mitogenomic phylogeny supports the position of suborder Brachycnemina as a clade within Macrocnemina coinciding with previous works (Poliseno *et al*., 2020). Therefore, Brachycnemina represents a paraphyletic group, with very high support both according to the Bayesian tree and the maximum-likelihood tree.

Even though the genome sequencing dataset of *Paleozoanthus reticulatus*, a specimen collected in 1982 (Table 1), showed signs of coverage issues with unreliable estimates of genome size, several mitochondrial genes could be retrieved from the sequencing data of this specimen. The specimen of *P. reticulatus* examined in this study is the only one reported since the species' original description in 1924 (Kise *et al.,* 2022), and its phylogenetic position within the family Epizoanthidae is has been unclear (Kise *et al.,* 2022). Although *Paleozoanthus* is associated with the gastropod genus *Granulifusus,* similar to *Epizoanthus protoporos* (Kise *et al.,* 2022), our molecular data suggest these species are not closely related. However, it has been previously suggested that this species might correspond to genus *Terrazoanthus,* in family *Hydrozoanthidae*, based on morphological features (Low *et al.,* 2016). Interestingly, the present phylogenetic reconstruction placed *Paleozoanthus reticulatus* within genus *Hydrozoanthus,* which belongs to the same family as *Terrazoanthus* (Kise *et al.,* 2019), Hydrozoanthidae. The phylogenetic placement of *Paleozoanthus reticulatus* within Hydrozoanthidae implies a previously undetected origin of symbioses with gastropods as members of this family are generally associated with hydroids, octocorals, or bare substrate, while mollusc-associated zoantharians had only been confirmed until now from family Epizoanthidae (Kise *et al.,* 2022,2023b). To clarify the phylogenetic position of *Paleozoanthus reticulatus,* including sequences of *Terrazoanthus* and other members of Hydrozoanthidae in future phylogenetic analyses is needed.

The present phylogeny also shows evidence of loss of symbiosis with Symbiodiniaceae within the family Sphenopidae, as the azooxanthellate species *Palythoa mizigama* and *Sphenopus marsupialis* were placed on internal branches within the primarily zooxanthellate genus *Palythoa.* This situation has been highlighted in previous phylogenies (Dudoit *et al.,* 2021) and it has been suggested that the loss of photosymbiosis may even have occurred twice (Irei *et al.,* 2015). However, samples from other azooxanthellate species of this family, *Palythoa umbrosa* are required to better clarify this point on the evolutionary history of photosymbiosis in Sphenopidae.

Genome size of zoantharians and the role of the repeatome in their dynamics

This study presents the first genome size measurements for zoantharians. Many estimates of genome sizes across the order Zoantharia were within expected measures for most cnidarians, namely between 500 Mbp and 700 Mbp (Adachi *et al.,* 2017). Among several genera of the order Zoantharia, genome sizes were found to overlap in their range (Fig. 2). For example, both genera *Zoanthus* and *Hydrozoanthus* included species with genome sizes of ~350 Mbp and 500 Mbp. It is possible that this pattern reflects intraspecific variations; zoantharian species may have retained genome sizes constrained in a similar range yet exhibit fluctuations within this range. Large intraspecific variations have been documented in invertebrates, as in the extreme case of snapping shrimps, in which disparities up to 6 Gbp have been observed within one species (Jeffery *et al.,* 2016). However, regarding cnidarians, the current knowledge points toward very narrow intervals; genome sizes are only known to vary up to 50 Mbp within jellyfish species *Sanderia malayensis* and *Rhopilema esculentum* (M.D. Santander, 2020, unpublished data) and less than 10 Mbp in anthozoans (Adachi *et al.*, 2017). Alternatively, it seems more likely that different zoantharian groups have undergone complex evolutionary dynamic processes resulting in interspecific genome size disparities of similar amplitudes.

The present results suggest that repeated elements, and in particular transposable elements, are involved in genome size dynamics of zoantharians, explaining at least partly the variations observed. Indeed, observed genome sizes were strongly correlated to the respective percentages of repeated and transposable elements (Fig. 6A and 6B). The paths to genome reduction or expansion are often the result of several processes, including transposable element activity or whole-genome duplication, which go in concert with changes in gene composition, genome structure and gene expression (Martín-Durán *et al.,* 2020). Other lines of evidence are required to fully understand the processes surrounding genome size variations in zoantharians, in particular from species of *Palythoa* and *Zoanthus*, as these genera show signs of hybridization (Reimer *et al.* 2007; Mizuyama *et al.,* 2018). However, the present results offer further insights into the contribution to genome size of various repeated elements. Similar to what Blommaert *et al.* (2019) observed with rotifers, a diversity of repeated elements was found in the repeatome of zoantharians (Fig. 2). The annotation of repeated elements was challenging, as up to 80% of identified repeats could not be successfully annotated by dnaPipeTE (Fig. 2). Due to the difficulty of repeated element assembly and annotation, unclassified elements are expected. Although in some insect groups, unclassified elements only account for ~10% of the total genome (Goubert *et al.,* 2015; Talla *et al.,* 2017), a study spanning several orders of Arthropoda showed a similar situation to our research, with more than 75% of repeats unclassified in some cases (Petersen *et al.* 2019). The number of unannotated repeats has also reached very high proportions in other cnidarians (Xia *et al.,* 2020). Such results may reflect the scarce number of repeat references from cnidarians in databases, calling for more efforts in characterizing repeatomes of cnidarians. Additionally, the use of short-read sequencing

566 may have contributed to the large amounts of unclassified repeats. However, annotation
567 is likely the main explanation, as our assemblies' N50 and contig numbers (Table S8)
568 were comparable to or better than those presented by the developers of dnaPipeTE
569 (Goubert *et al.,* 2015), who obtained significantly fewer unclassified elements.

570 Although we obtained large proportions of unclassified repeats in the dnaPipeTE
571 analyses, the clustering and repeat annotation performed via RepeatExplorer2
572 suggested that they may be partly represented by satellite elements (Fig.5, Fig. S3).
573 Indeed, they accounted for ~30% of the annotated elements in the comparative analysis
574 (Table S7), yet they were almost absent from annotations via dnaPipeTE (Fig.2).
575 Conversely, numerous mobile elements could not be annotated from RepeatExplorer2.
576 While this partly reflects the different sensitivities of the two pipelines and the databases
577 that they use, the consistently high amounts of unclassified repeats in zoantharians
578 highlight that much remains to be discovered with regards to their genomes. More
579 efforts into assembling and characterizing their repeatomes will surely reveal interesting
580 elements. Indeed, the percentages of unclassified repeat categories were found to be
581 correlated to genome size (Fig. 6D), suggesting that elements with significance for
582 genome size dynamics are contained among unclassified repeats.

583 Of the repeated elements that could be annotated, the most abundant classes were
584 DNA transposons and LINEs elements. These results are in line with previous studies
585 on the repeated DNA content of several cnidarians, where these two classes were also
586 observed to be the most abundant (Xia *et al.,* 2020).

587 The literature on the roles of repeated elements in genome sizes has largely focused on
588 cases displaying extreme genome size variations. In these situations, dramatic changes
589 of genome sizes in association with a single specific repeated class have been reported.
590 Notably, the class of repeated elements involved varies between taxa; in larvaceans
591 SINEs elements appear to drive genome size increases (Naville *et al.*, 2019), while
592 satellites and helitrons were the main contributors in migratory locusts (Shah *et al.,*
593 2020). In *Hydra,* LINEs elements have had a major expansion event leading to dramatic
594 genome size increase in the subgroup of brown *Hydra* (Wong *et al.,* 2019*).* Although
595 different repeated elements are clearly involved in genome size dynamics in different
596 groups, the degree of variation between taxa is not well understood. The present
597 dataset offers insights in this question by adding to the knowledge of repeated elements
598 in Cnidaria. Genome size variations in zoantharians do not appear to be as important as
599 in *Hydra*, but still reach a maximum variation of 2.4X fold, between congeners *Palythoa*
600 *tuberculosa* and *P. mizigama.* However, it is notable that LINEs elements – the class
601 responsible for genome size expansion in *Hydra* – were consistently one of the most
602 abundant in our dataset (Fig. 2B), and that a significant correlation between this class
603 and genome size was detected (Fig. 6C). Furthermore, the repeat landscapes of most
604 species showed a high number of LINEs elements with low divergence (Fig. 4). Such
605 patterns have been interpreted as a sign of recent TE activity; TE copies in the
606 genomes accumulate at a faster rate than mutations in their sequences (Goubert *et al.,*

2015). Among them, two subfamilies appeared to be particularly abundant; namely LINE/L2 and Penelope elements (Fig. 4, Fig. S2). LINE/L2 were also one of the most abundant elements in the brown *Hydra* group (Wong *et al.,* 2019) as well as *Aurelia* jellyfish (Khalturin *et al.,* 2019). Therefore, this set of evidence suggests that the activity of various LINEs elements may have led to increased genome sizes in zoantharians, and potentially may have done so across Cnidaria. In parallel with their effects on genome size, LINE/L2 elements may have impacted the evolution and functioning of zoantharians. Indeed, their role in the regulatory networks of housekeeping genes through the activity of LINE/L2-derived miRNAs have been demonstrated in humans (Petri *et al.,* 2019).

Another seemingly important group of repeated elements in zoantharians are satellite elements, which represented the most numerous and largest clusters in the RepeatExplorer2 analysis (Fig. 6, Fig. S3). The comparative analysis performed via RepeatExplorer2 revealed instances of species-specific differentially expanded clusters. Closely related species (such as the pairs *H. antumbrosus* and *H. tunicans, B. catenularis* and *B. puertoricense*, *Z. solanderi* and *Z. gigantus,* Fig. 1) showed almost identical amplified clusters (Fig. 5, Fig. S3). On the contrary, species of those same genera but that branched earlier in the phylogenetic tree (Fig. 1) such as *H. sils,* or species that were simply more divergent, such as *Z. pulchellus* and *Z. sociatus*, showed unique cluster amplifications. These instances confirm that the species pairs mentioned above are very closely related, but also indicate that different satellite elements are amplified in the genomes of different species over the course of their evolution. Furthermore, this phylogenetic pattern is consistent with genome size dynamics. Indeed, several species that display large satellite elements clusters have larger genomes compared to other species of their group (*H. sils, Z. pulchellus*, and *S. marsupialis,* Fig. 2). In the migratory locust, expansion of satellite elements in the largest genomes were observed (Shah *et al.,* 2020). These authors suggested that rather of a causal relationship, the proliferation of satellites could be a consequence of genome expansion, as a mean to protect centromeric and telomeric chromosome regions after genome enlargement from transposable elements (Shah *et al.,* 2020). Considering their occurrence in species that have diverged for a long period of time, this may also possibly be the case in zoantharians. However, the largest genome detected in this study, that of *P. tuberculosa,* did not display such large cluster amplifications of satellite elements. Together with our results on LINEs and unclassified elements, and we conclude that the genome size patterns observed in zoantharians are likely the result of the activity of multiple groups of repeated elements.

Traits potentially associated with genome size and repeated DNA

Two main evolutionary theories have been proposed to explain the puzzling variations observed in genome sizes; one that focuses on neutral processes and one on selective processes (Blommaert *et al.,* 2020). In the first theory, the accumulation of DNA is considered to be a result of drift. The opposite theory suggests genome size is under

the influence of selective forces and may impact organismal traits. In particular, genome size has been correlated to body size and egg size (Naville *et al.,* 2019; Stelzer *et al.,* 2021), giving support to the nucleotypic hypothesis that proposes that genome size directly impacts phenotype by an effect on cell volume. However other traits have been suggested to potentially be impacted by genome sizes, including geographical distribution (Leinaas *et al.,* 2016), habitat (Paule *et al.,* 2021) and effective population sizes (Lefébure *et al.,* 2017). Although we did not formally analyze variations of genome sizes with phenotypic or biogeographic characteristics, a comparison with the phylogeny of zoantharians hints at features that may be affected. Symbiosis with Symbiodiniaceae dinoflagellates is one of the most studied facets of cnidarian biology because of its importance in sustaining the life of reef-building cnidarians and the subtropical to tropical ecosystem they support. This interaction is endosymbiotic and has large influence on host metabolism at the cellular level (Davy *et al.,* 2012), which in line with the nucleotypic hypothesis would have the potential to negatively impact genome size (Adachi *et al.,* 2017). Because of this, a former investigation of genome sizes in Cnidaria attempted to find correlations between Symbiodiniaceae symbiosis and genome size (Adachi *et al.,* 2017), but did not observe any significant relationship. However, in *Hydra,* genome size expansion has been associated with a switch away from symbiotic lifestyle (Wong *et al.,* 2019). Indeed, the green hydra, with small genomes, maintains an obligate relationship with *Chlorella,* while symbiosis is not mandatory for strains of the brown *Hydra,* which have enlarged genomes (Ishikawa *et al.,* 2016; Wong *et al.,* 2019). In our study, on the other hand, a contrasting pattern was revealed between genome sizes and symbiosis in the group *Palythoa*. This genus comprised the largest genome size variation observed in all zoantharians – a 2.4x fold variation – between species with different symbiotic lifestyles. The maximum genome size was in zooxanthellate *P. tuberculosa* – 678 Mbp, while the minimum size was in azooxanthellate *P. mizigama,* with 286 Mbp (Fig. 2). This makes *P. mizigama* within the range of the smallest cnidarian genomes recorded, that of *Sanderia malayensis* with a C-value of 0.26 pg, or about 250 Mbp (Adachi *et al.,* 2017). Since macrocnemic zoantharians have similar ranges of genome sizes to brachycnemic zooxanthellate *Zoanthus* spp., it seems that the switch to a Symbiodiniaceae-associated lifestyle did not impact genome size. However, based on the *Palythoa r*esults, the loss of this relationship may be associated with smaller genome sizes. It can be hypothesized that the activity or loss of repeated DNA accompanying genome size reduction of *P. mizigama* may have caused some genomic rearrangements impacting functions linked to symbiosis. As the loss of symbiosis may have occurred several times in *Palythoa* (Irei *et al.,* 2015), the rapid activity and movement of TE may be partly behind the apparent "switching on and off" of symbiosis in this group. Conversely, it is apparent that *P.tuberculosa* experienced genome enlargement. Following the reasoning of the nucleotypic hypothesis, genome sizes can be expected to be smaller in the case of a symbiotic organism, due to the symbiont effect on cell volume and metabolism (Adachi *et al.,* 2017). However, symbiotic species may be subject to genome size increase through horizontal transfer and activity of transposable

692 elements of their symbiotic counterpart. Although there is no documented evidence of
693 TE transfer between Symbiodiniacae and hosts, transposable elements transcripts in
694 *Symbiodinium* have been shown to be upregulated in situations of environmental stress
695 (Chen *et al.,* 2018). Such event may have contributed to large genome size observed in
696 the case of *P. tuberculosa.*

697 Alternatively, potential past event of hybridization may have contributed. Hybridization is
698 known to have occurred in zoantharians (Reimer *et al.*, 2007) including genus *Palythoa*
699 (Mizuyama *et al*., 2018). Hybridization is thought to potentially trigger the activation of
700 TE, leading to their accumulation in the hybrid genome (Baack, Whitney & Rieseberg,
701 2005; Hénault *et al.,* 2020). This may have promoted species reproductive isolation as
702 the increased transposition activity may have deleterious effects and cause sterility of
703 the hybrids of two divergent populations (Dion-Côté *et al.,* 2014; Serrato-Capuchina *et*
704 *al.,* 2018), and may have contributed to the evolution of *P. tuberculosa, P.* sp. yoron, *P.*
705 *mutuki* and *P.* aff. *Mutuki* (Mizuyama *et al.,* 2018). Multiple aspects of zoantharian
706 biology may be associated with genome size variations and transposable elements
707 activity. To further understand the potential relationships between them, genome
708 assemblies and estimates of genome sizes for other *Palythoa* species are necessary.

709

710 ## Conclusions

711 In this paper, we explored the relationships between phylogeny, genome size variations,
712 and the repetitive elements composition of a scarcely studied group of cnidarians. Our
713 results show that genome sizes observed in zoantharians are likely the product of
714 complex historical dynamics of the repeatome. We found a high number of unknown
715 repeats with potential implications in genome size. Recent expansion events of LINEs,
716 DNA and satellite elements were identified in multiple species, raising questions on the
717 role of these elements in genome evolution of cnidarians and the consequences of their
718 activity. Until now no information was available for zoantharian genome sizes, and we
719 here present such information for 18 specimens from five of the nine zoantharian
720 families. This research demonstrates the power of next-generation sequencing projects
721 aimed at understudied taxa, allowing a rapid increase in our basic understanding of
722 such poorly studied groups. This sequencing project also allowed us to clarify the
723 phylogenetic position of *Paleozoanthus* via analyses of an old specimen; such work
724 could very likely not have been performed utilizing traditional genetic methods. Finally,
725 as there are notable questions related to the ecology, symbioses, development and
726 evolution of zoantharians, the genome data and repeatome characteristics presented
727 here will serve as important baseline data to investigate such questions in future
728 genomic projects.

729

References

Adachi, K., Miyake, H., Kuramochi, T., Mizusawa, K., & Okumura, S. (2017). Genome size distribution in phylum Cnidaria. *Fisheries Science*, *83*(1), 107–112.

Akaike,H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, 267–281.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Aratake, S., Taira, Y., Fujii, T., Roy, M. C., Reimer, J. D., Yamazaki, T., & Jenke-Kodama, H. (2016). Distribution of palytoxin in coral reef organisms living in close proximity to an aggregation of *Palythoa tuberculosa*. *Toxicon*, *111*, 86–90.

Baack, E.J., Whitney, K.D. & Rieseberg, L.H. (2005). Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *New Phytologist*, *167*(2):623-30.

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, *6*(1), 11.

Barrett, N. J., Hogan, R. I., Allcock, A. L., Molodtsova, T., Hopkins, K., Wheeler, A. J., & Yesson, C. (2020). Phylogenetics and mitogenome organization in black corals (Anthozoa: Hexacorallia: Antipatharia): An order-wide survey inferred from complete mitochondrial genomes. *Frontiers in Marine Science*, *7*, 440.

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

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-20.

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, *19*(1), 199.

Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner* (LBNL-7065E). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). https://sourceforge.net/projects/bbmap/

Chen, J., Cui, G., Wang, X., Liew, Y.J. & Arand, M. (2018). Recent expansion of heat-activated retrotransposons in the coral symbiont *Symbiodinium microadriaticum*. *The ISME Journal, 12*, 639–643.

Chi, S. I., & Johansen, S. D. (2017). Zoantharian mitochondrial genomes contain unique complex group I introns and highly conserved intergenic regions. *Gene*, *628*, 24–31.

Davy, S. K., Allemand, D., & Weis, V. M. (2012). Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiology and Molecular Biology Reviews*, *76*(2), 229–261.

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data, *Nucleic Acids Research*, *45*(4), e18.

770 Dion-Côté, A.M., Renaut. S-, Normandeau. E. & Bernatchez, L. (2014). RNA-seq reveals
771 transcriptomic shock involving transposable elements reactivation in hybrids of young
772 lake whitefish species. *Molecular Biology and Evolution, 31*(5):1188-99.
773 Dudoit, A., Santos, M. E. A., Reimer, J. D., & Toonen, R. J. (2021). Phylogenomics of
774 *Palythoa* (Hexacorallia: Zoantharia): probing species boundaries in a globally distributed
775 genus. *Coral Reefs*. doi:10.1007/s00338021-02128-4.
776 Elliott, T. A., & Gregory, T. R. (2015). What's in a genome? The C-value enigma and the
777 evolution of eukaryotic genome content. *Philosophical Transactions of the Royal*
778 *Society of London. Series B, Biological Sciences*, *370*(1678), 20140331.
779 Goubert, C., Modolo, L., Vieira, C., ValienteMoro, C., Mavingui, P., & Boulesteix, M.
780 (2015). De novo assembly and annotation of the Asian tiger mosquito (*Aedes*
781 *albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative
782 analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biology and Evolution*,
783 *7*(4), 1192–1205.
784 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I.,
785 Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N.,
786 Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K.,
787 Friedman, N., & Regev, A. (2011). Trinity: Reconstructing a full-length transcriptome
788 without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652.
789 Gregory, T. R. (2005). The C-value enigma in plants and animals: A review of parallels
790 and an appeal for partnership. *Annals of Botany*, *95*(1), 133–146.
791 Hahn, M. W., & Wray, G. A. (2002). The g-value paradox. *Evolution & Development*, *4*(2),
792 73–75.
793 Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a
794 molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160–174.
795 Hénault, M., Souhir, M., Charron, G. & Landry, C.R. (2020). The effect of hybridization on
796 transposable element accumulation in an undomesticated fungal species. *eLife.*
797 9:e60474.
798 Hirose, M., Obuchi, M., Hirose, E. & Reimer, J.D, (2011). Timing of spawning and early
799 development of *Palythoa tuberculosa* (Anthozoa, Zoantharia, Sphenopidae) in Okinawa,
800 Japan. *The Biological Bulletin, 220(1),* 23-31.
801 Hotaling, S., Kelley, J. L., & Frandsen, P. B. (2021). Toward a genome sequence for
802 every animal: Where are we now? *Proceedings of the National Academy of Sciences*,
803 *118*(52).
804 Irei, Y., Sinniger, F., & Reimer, J. D. (2015). Descriptions of two azooxanthellate *Palythoa*
805 species (Subclass Hexacorallia, Order Zoantharia) from the Ryukyu Archipelago,
806 southern Japan. *ZooKeys*, *478*, 1–26.
807 Ishikawa, M., Yuyama, I., Shimizu, H., Nozawa, M., Ikeo, K., & Gojobori, T. (2016).
808 Different endosymbiotic interactions in two hydra species reflect the evolutionary history
809 of endosymbiosis. *Genome Biology and Evolution*, *8*(7), 2155–2163.

Jeffery, N. W., Hultgren, K., Chak, S. T. C., Gregory, T. R., & Rubenstein, D. R. (2016). Patterns of genome size variation in snapping shrimp. *Genome*, *59*(6), 393–402.

Johansen, S.D., Chi, S.I., Dubin, A., & Jørgensen, T.E. (2021) The mitochondrial genome of the sea anemone *Stichodactyla haddoni* reveals catalytic introns, insertion-like element, and unexpected phylogeny. *Life*, *11(5)*, 402.

Kayal, E., Bentlage, B., Pankey, S. 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*.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649.

Khalturin, K., Shinzato, C., Khalturina, M., Hamada, M., Fujie, M., Koyanagi, R., Kanda, M., Goto, H., Anton-Erxleben, F., Toyokawa, M., Toshino, S., & Satoh, N. (2019). Medusozoan genomes inform the evolution of the jellyfish body plan. *Nature Ecology and Evolution, 3*, 811–822.

Kise, H., Maeda, T., & Reimer, J. D. (2019). A phylogeny and the evolution of epizoism within the family *Hydrozoanthidae* with description of a new genus and two new species. *Molecular Phylogenetics and Evolution*, *130*, 304–314.

Kise, H., Obuchi, M., & Reimer, J. D. (2021). A new *Antipathozoanthus* species (Cnidaria, Hexacorallia, Zoantharia) from the northwest Pacific Ocean. *ZooKeys*, *1040*, 49–64.

Kise, H., Moritaki, T., Iguchi, A., & Reimer, J.D. (2022). Epizoanthidae (Hexacorallia: Zoantharia) associated with *Granulifusus* gastropods (Neogastropoda: Fasciolariidae) from the Indo-West Pacific. *Organisms Diversity and Evolution*.

Kise, H., Reimer, J.D & Pirro S. (2023). The Complete Genome Sequences of 7 Species of *Epizoanthus* (Epizoanthidae, Zoantharia, Hexacorallia, Cnidaria). *Biodiversity Genomes*, March.

Kise, H., Santos, M.E.A., Fourreau, C.J.L., Iguchi, A., Goto. R. & Reimer, J.D. (2023). Evolutionary patterns of host switching, lifestyle mode, and the diversification history in symbiotic zoantharians. *Molecular Phylogenetics and Evolution*, *11*, 182:107732.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RaxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*, 1870–1874.

Lefébure, T., Morvan, C., Malard, F., François, C., Konecny-Dupré, L., Guéguen, L. Weiss-Gayet, M., Seguin-Orlando, A., Ermini, L., Der Sarkissian, C., Charrier, N.P.,

849 Eme, D., Mermillod-Blondin, F., Duret, L., Vieira, C., Orlando, L. & Douady, C. J. (2017).
850 Less effective selection leads to larger genomes. *Genome Research*, 27(6), 1016-1028.
851 Leinaas, H. P., Jalal, M., Gabrielsen, T. M., & Hessen, D. O. (2016). Inter- and
852 intraspecific variation in body- and genome size in calanoid copepods from temperate
853 and arctic waters. *Ecology and Evolution*, *6*(16), 5585–5595.
854 Lin, M.-F., Kitahara, M. V., Luo, H., Tracey, D., Geller, J., Fukami, H., Miller, D. J., &
855 Chen, C. A. (2014). Mitochondrial genome rearrangements in the
856 Scleractinia/Corallimorpharia Complex: Implications for coral phylogeny. *Genome*
857 *Biology and Evolution*, *6*(5), 1086–1095.
858 Low, M. E. Y., Sinniger, F., & Reimer, J. D. (2016). The order Zoantharia Rafinesque,
859 1815 (Cnidaria, Anthozoa: Hexacorallia): supraspecific classification and nomenclature.
860 *ZooKeys*, *641*, 1–80.
861 Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel
862 counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770.
863 Martín-Durán, J.M., Vellutini, B.C., Marlétaz, F., Cetrangolo, V., Cvetesic, N., Thiel, D.,
864 Henriet, S., Grau-Bové, X., Carrillo-Baltodano, A.M., Gu, W., Kerbl, A., Marquez, Y.,
865 Bekkouche, N., Chourrout, D., Gómez-Skarmeta, J.L., Irimia, M., Lenhard, B., Worsaae,
866 K., & Hejnol, A. (2021). Conservative route to genome compaction in a miniature
867 annelid. *Nature Ecology and Evolution*, *5*, 231–242.
868 Metzger, M. J., Paynter, A. N., Siddall, M. E., & Goff, S. P. (2018). Horizontal transfer of
869 retrotransposons between bivalves and other aquatic species of multiple phyla.
870 *Proceedings of the National Academy of Sciences*, *115*(18), E4227–E4235.
871 Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., Shaw, P.D. &
872 Marshall, D. (2013). Using Tablet for visual exploration of second-generation
873 sequencing data. *Briefings in Bioinformatics, 14*(2), 193-202.
874 Mizuyama, M., Masucci, G. D., & Reimer, J. D. (2018). Speciation among sympatric
875 lineages in the genus *Palythoa* (Cnidaria: Anthozoa: Zoantharia) revealed by
876 morphological comparison, phylogenetic analyses and investigation of spawning period.
877 *PeerJ*, *6*, e5132.
878 Montenegro, J., Sinniger, F., & Reimer, J. D. (2015). Unexpected diversity and new
879 species in the sponge-*Parazoanthidae* association in southern Japan. *Molecular*
880 *Phylogenetics and Evolution*, *89*, 73–90.
881 Naville, M., Henriet, S., Warren, I., Sumic, S., Reeve, M., Volff, J-N., & Chourrout, D.
882 (2019). Massive changes of genome size driven by expansions of non-autonomous
883 transposable elements. *Current Biology*, *29*(7), 1161–1168.
884 Novák, P., Neumann, P., & Macas, J. 2020. Global analysis of repetitive DNA from
885 unassembled sequence reads using RepeatExplorer2. *Nature Protocols*, *15*(11), 3745–
886 3776.

887 Pantzartzi, C. N., Pergner, J., & Kozmik, Z. (2018). The role of transposable elements in
888 functional evolution of amphioxus genome: The case of opsin gene family. *Scientific*
889 *Reports*, *8*(1), 2506.
890 Paule, J., von Döhren, J., Sagorny, C., & Nilsson, M. A. (2021). Genome size dynamics
891 in marine ribbon worms (Nemertea, Spiralia). *Genes*, *12*(9), 1347.
892 Petersen, M., Armisén, D., Gibbs, R. A., Hering, L., Khila, A., Mayer, G., Richards, S.,
893 Niehuis, O., & Misof, B. (2019). Diversity and evolution of the transposable element
894 repertoire in arthropods with particular reference to insects. *BMC Ecology and Evolution*,
895 *19*(1), 11.
896 Petri, R., Brattås, P. L., Sharma, Y., Jönsson, M. E., Pircs, K., Bengzon, J., & Jakobsson,
897 J. (2019). LINE-2 transposable elements are a source of functional human microRNAs
898 and target sites. *PLoS Genetics*, *15*(3), e1008036.
899 Poliseno, A., Santos, M.E.A., Kise, H., MacDonald, B., Quattrini, A.M., McFadden, C.S.,
900 & Reimer, J.D. (2020). Evolutionary implications of analyses of complete mitochondrial
901 genomes across order Zoantharia (Cnidaria: Hexacorallia). *Journal of Zoological*
902 *Systematics and Evolutionary Research, 58*(4), 858-868.
903 Quattrini, A. M., Rodríguez, E., Faircloth, B. C., Cowman, P. F., Brugler, M. R., Farfan, G.
904 A., Hellberg, M. E., Kitahara, M. V., Morrison, C. L., Paz-García, D. A., Reimer, J. D., &
905 McFadden, C. S. (2020). Palaeoclimate ocean conditions shaped the evolution of corals
906 and their skeletons through deep time. *Nature Ecology and Evolution*, *4*(11), 1531–1538.
907
908 R Core Team. (2021). R: A language and environment for statistical computing. *R*
909 *Foundation for Statistical Computing, Vienna, Austria.* https://www.R-project.org/
910 Reimer, J.D., Takishita, K., Ono, S., Tsukahara, J., & Maruyama, T. (2007). Molecular
911 evidence suggesting interspecific hybridization in *Zoanthus spp*. (Anthozoa:
912 Hexacorallia) *Zoological Science*, *24*, 346–359.
913 Ronquist F., & Huelsenbeck J. (2003). MrBayes 3: Bayesian phylogenetic inference
914 under mixed models. *Bioinformatics*, *19*, 1572–1574.
915 Santos, M.E.A., Wirtz, P., Montenegro, J., Kise, H., López, C., Brown, J., & Reimer, J.
916 (2019). Diversity of Saint Helena Island and zoogeography of zoantharians in the
917 Atlantic Ocean: Jigsaw falling into place *Systematics and Biodiversity*, *17*(2), 165–178.
918 Santos, M. E. A., Kise, H., Fourreau, C.J.L, Poliseno, A., Pirro, S., & Reimer, J.D. (2023).
919 The complete genome sequences of 13 species of Brachycnemina (Cnidaria,
920 Hexacorallia, Anthozoa, Zoantharia). *Biodiversity Genomes*, March.
921 Stampar, S.N., Broe, M.B., Macrander, J., Reitzel, A.M., Brugler, M.R. & Daly, M. (2019).
922 Linear mitochondrial genome in Anthozoa (Cnidaria): a case study in Ceriantharia.
923 *Scientific Reports, 9, 6094*.
924 Serrato-Capuchina, A., & Matute, D. R. (2018). The role of transposable elements in
925 speciation. *Genes*, *9*(5), 254.

926 Sinniger, F., Montoya-Burgos, J.I., Chevaldonné, P. & Pawloski, J. (2005). Phylogeny of
927 the order Zoantharia (Anthozoa, Hexacorallia) based on the mitochondrial ribosomal
928 genes. *Marine Biology, 147,* 1121-1128.
929 Sinniger, F., Chevaldonné, P., & Pawlowski, J. W. (2007). Mitochondrial genome of
930 *Savalia savaglia* (Cnidaria, Hexacorallia) and early metazoan phylogeny. *Journal of*
931 *Molecular Evolution*, *64*(2), 196.
932 Stelzer, C.-P., Pichler, M., & Hatheuer, A. (2021). Linking genome size variation to
933 population phenotypic variation within the rotifer, *Brachionus asplanchnoidis*.
934 *Communications Biology*, *4*(1), 1–8.
935 Swift, H. (1950). The constancy of desoxyribose nucleic acid in plant nuclei. *Proceedings*
936 *of the National Academy of Sciences of the United States of America*, *36*(11), 643–654.
937 Talla, V., Suh, A., Kalsoom, F., Dincă, V., Vila, R., Friberg, M., Wiklund, C., & Backström,
938 N. (2017). Rapid increase in genome size as a consequence of transposable element
939 hyperactivity in wood-white (*Leptidea*) Butterflies. *Genome Biology and Evolution*, *9*(10),
940 2491–2505.
941 Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are
942 strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution*,
943 *9*(4), 678–687.
944 Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA
945 sequences. *Lectures on Mathematics in the Life Sciences, 17*, 57–86.
946 Van't Hof, A.E., Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N.,
947 Darby, A.C. & Saccheri. I.J. (2016). The industrial melanism mutation in British
948 peppered moths is a transposable element. *Nature*, *2;534*(7605), 102-5.
949 Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: Concatenation software
950 for the fast assembly of multi-gene datasets with character set and codon information.
951 *Cladistics*, *27*(2), 171–180.
952 Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New
953 York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org
954 Wicker, T., Sabot, F. Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub B., Flavell, A.,
955 Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P. & Schulman, A.H. (2007).
956 A unified classification system for eukaryotic transposable elements. *Nature Reviews*
957 *Genetics*, *8*, 973-982.
958 Wong, W. Y., Simakov, O., Bridge, D. M., Cartwright, P., Bellantuono, A. J., Kuhn, A.,
959 Holstein, T. W., David, C. N., Steele, R. E., & Martínez, D. E. (2019). Expansion of a
960 single transposable element family is associated with genome-size increase and
961 radiation in the genus *Hydra*. *Proceedings of the National Academy of Sciences*,
962 *116*(46), 22915–22917.
963 Xia, W., Li, H., Cheng, W., Li, H., Mi, Y., Gou, X., & Liu, Y. (2020). High-quality genome
964 assembly of *Chrysaora quinquecirrha* provides insights into the adaptive evolution of
965 jellyfish. *Frontiers in Genetics*, *11*, 535.

966 Yang, S.-Y., Bourgeois, C., Ashworth, C. D., & Reimer, J. D. (2013). *Palythoa* zoanthid
967 'barrens' in Okinawa: Examination of possible environmental causes. *Zoological Studies*,
968 *52*(1), 39.
969 Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Current Protocols in*
970 *Bioinformatics*, *69*:e96.
971

972 Acknowledgements

973 CJLF is grateful to MEXT (Ministry of Education, Culture, Sports, Science and
974 Technology of Japan) for a scholarship and to a University of the Ryukyus ORCIDS
975 grant for training workshop funding. We are grateful to NIG (National Institute of
976 Genetics, Japan) for allowing us to use their supercomputer cluster for bioinformatic
977 analyses, and in particular Tomohiro Hirai for their assistance in trouble shooting. We
978 are also grateful to the ELIXIR-CZ project, which provided additional computational
979 resources for the Repeat Explorer 2 analysis. We are thankful to Assoc. Prof. Takashi
980 Nakamura, for his revision of a previous version of this manuscript and providing
981 comments on the discussion. MSD was supported by Coordenação de
982 Aperfeiçoamento de Pessoal de Nível Superior 88882.377420/2019-01 and MMM by
983 Fundação de Amparo à Pesquisa do Estado de São Paulo FAPESP 2016/04560-9 and
984 FAPESP BEPE 2017/25907-0.

Table 1(on next page)

Table 1. Specimen investigated for their next-generation sequencing data in this study, with corresponding information.

Species	Original specimen number	Sampling date	Location	Depth (m)	GPS coordinates	Nb Reads	Accession	SRA number	Mitochondrial genome Accession
Antipathozoanthus remengesaui	2073168	27/06/2014	Onna, Okinawa, Japan	15	N 26°26'20", E 127°47'7"	67582218	PRJNA598175	SRR11206917	OQ759540
Antipathozoantus obscurus	2073167	14/08/2014	Bise, Motobu, Okinawa, Japan	35	N 26°42'34", E 127°52'49"	63225999	PRJNA598176	SRR11206396	OQ759541
Bergia catenularis	JDR170610-4-30	17/06/2017	St Michiel, Curacao	20	12°08'05“N, 69°00'00“W	91954038	PRJNA662740	SRR12621188	OQ740722
Bergia puertoricense	JDR170612-7-45	17/06/2017	Marie Pampoen, Curacao	31	12°05'02“N, 68°54'01“W	79593063	PRJNA662980	SRR12626556	OQ740723
Epizoanthus illoricatus	328561	27/02/2015	Blue Corner, Palau	31	N 7°8'12", E 134°13'16"	66320841	PRJNA598181	SRR11206918	OQ740724
Epizoanthus planus	283861	13/03/2008	Hirajisone, Nagasaki, Japan	280	N 32°15'82", E °129'16"	31291805	PRJNA676135	SRR13036371	NA
Epizoanthus ramosus	760129	19/09/2019	Minnajima, Okinawa, Japan	210	N 26°38'41",E 127°47'34"	62798707	PRJNA598173	SRR11207078	OQ740725
Epizoanthus rinbou	1351471	13/03/2008	Hirajisone, Nagasaki, Japan	280	N 32°15'82", E °129'16"	71661186	PRJNA662986	SRR12626621	OQ740726

Epizoanthus scotinus	283864	12/07/2018	Sgaan Kinghlas-Bowie Seamount, Canada	86	N 53°18'5", E 135°40'36"	87383374	PRJNA662773	SRR12621595	OQ740727
Hydrozanthus tunicans	JDR170609-2-7	09/06/2017	Water Factory, Curacao	30	12°06'03“N, 68°57'01“W	76797899	PRJNA645597	SRR12201157	OQ740731
Hydrozoanthus antumbrosus	JDR191030-2-2	30/10/2019	It's Pretty Rough, Bonaire	25	N 12°04'882'' , W 068°13'926''	8939118	PRJNA662983	SRR12626620	OQ740728
Hydrozoanthus gracilis	283869	27/06/2019	Cape Ose, Izu, Shizuoka, Japan	30	N 35°01'50", E 138°47'12"	74323024	PRJNA662988	SRR12626630	OQ740730
Hydrozoanthus sils	1320743	13/09/2015	Ngardmau, Palau	24	N 7°26’17.0’’ , E 134°36’49.4’’	77513745	PRJNA662735	SRR12621138	OQ740729
Paleozoanthus reticulatus	581563	15/07/1982	Transkei, South Africa	100	NA	49040454	PRJNA622546	SRR12621205	OQ843460 (contig), OQ848443 (COI)
Palythoa carribaeorum	180802-2.14	02/08/2018	Madagascar Reef, Sisal, Yucatan, Mexico	1	N 21°26’17” , W 90°16’39”	63858955	PRJNA598184	SRR11206360	OQ785262
Palythoa grandiflora	170419-65	19/04/2017	Puerto Viejo, Limon, Puerto Limon, Costa Rica	1	N 9°39'33", W 82°45'12"	65100764	PRJNA598185	SRR11206528	OQ785263
Palythoa grandis	JDR170613-	13/06/2017	Tugboat,	12	12°04'00“N,	64824964	PRJNA580275	SRR12621764	OQ785264

	10-62		Curacao		68°51'04“W				
Palythoa heliodiscus	JDR191205-1-1	05/12/2019	Mizugama, Kadena, Okinawa, Japan	8	N 26°21'33", E 127°44'17"	63700889	PRJNA598194	SRR11206406	OQ785265
Palythoa mizigama	A10	08/07/2018	Mizugama, Kadena, Okinawa, Japan	5	N 26°21'33", E 127°44'17"	55318416	PRJNA957836	SRR24234327	OQ785267
Palythoa mutuki	OKW13	13/11/2017	Adan Beach, Oku, Okinawa, Japan	3	N 26°49'21", E 128°18'45"	66775959	PRJNA598187	SRR11206913	OQ789241
Palythoa tuberculosa		05/02/2019	Mizugama, Kadena, Okinawa, Japan	7	N 26°21'33", E 127°44'17"	302007118, 629888728	PRJNA946699	SRR23916682	OQ843460
Parazoanthus swiftii	JDR170609-2-6	09/06/2017	Water Factory, Curacao	21	12°06'03“N, 68°57'01“W	77782303	PRJNA662982	SRR12626618	OQ785266
Parazoantus darwini	461	Mar-07	Espanola, Galapagos, Equador	11	1°21'52" S 89°38'07"W	38229398	PRJNA662981	SRR12626557	OQ785268
Sphenopus marsupialis	291061	07/06/2019	off Nakijin, Okinawa, Japan		N 26°43’30”, E 127°56’47’’	82098096	PRJNA662993	SRR12626632	OQ759543
Umimayanthus	62JR	26/08/2010	North Directi	12	S 14°45’0	74274477	PRJNA662702	SRR12620700	NA

chanpuru			on Island, Queensland, Australia		3”S,E 145°30’ 43”				
Umimayanthus nakama	363JR	13/09/2006	Otsuki, Kochi, Japan	3	N 32°46’53”,E 132°40’09”	81231918	PRJNA645598	SRR12201158	OQ759542
Umimayanthus parasiticus	JDR170609-1-1	09/06/2017	Hilton, Curacao	31	12°07'14"N 68°58'12"W	85383046	PRJNA662764	SRR12621190	OQ785261
Zoanthus gigantus	JDR191205-1-2	05/12/2019	Mizugama, Kadena, Okinawa, Japan	15	N 26°21'49", E 127°46"21"	67178822	PRJNA598193	SRR11206404	OQ785260
Zoanthus pulchellus	JDR170613-9-56	13/06/2017	Director's Bay, Curacao	10	12°03'05“N, 68°51'03“W	67487982	PRJNA645596	SRR12201156	OQ759536
Zoanthus sociatus	170418-23	2017.04.18	Piuta Beach, Limon, Puerto Limon, Costa Rica	1	10°00'19.3"N 83°02'02.4"W	71698365	PRJNA598186	SRR11206569	OQ759538
Zoanthus solanderi	JDR170620-23-99	2017.06.20	Playa Jeremi, Curacao		12°19'45"N, 69°09'05"W	86269889	PRJNA662769	SRR12621302	OQ759539
Zoanthus sansibaricus	OKW21	2017.11.13	Adan Beach, Oku, Okinawa, Japan	3	26°49'21.8"N 128°18'45.9"E	66995947	PRJNA598188	SRR11206971	OQ759537

1

Table 2(on next page)

Table 2. Previously assembled mitochondrial genomes included in this study.

Species	Museum Voucher	Sampling location	Accession	Reference
Palythoa heliodiscus	MISE MS160525-33	Dongsha, Taiwan	MN873593	Poliseno et al. (2020)
Palythoa heliodiscus	NA	Aquarium trade	KY888673	Chi & Johansen (2017)
Palythoa mutuki	MISE JDR160604-44	Dongsha, Taiwan	MN873595	Poliseno et al. (2020)
Palythoa mizigama	MISE201705 MizugamaPmiz	Okinawa, Japan	MN873594	Poliseno et al. (2020)
Sphenopus marsupialis	MISE S8	Brunei	MN873598	Poliseno et al. (2020)
Zoanthus sansibaricus	NA	Okinawa, Japan	NC035578	Poliseno et al. (2020)
Zoanthus cf. *sociatus*	NA	Atlantic side, Panama	MN873600	Poliseno et al. (2020)
Zoanthus sociatus	MISE JDR150614-125	St Eustatius, The Netherlands	MN873601	Poliseno et al. (2020)
Zoanthus pulchellus	PAB-15-56	Atlantic side, Panama	MN873599	Poliseno et al. (2020)
Zoanthus sp.	PAN-23	Atlantic side, Panama	MN873602	Poliseno et al. (2020)
Epizoanthus illoricatus	MISE 140519	Okinawa, Japan	MN873588	Poliseno et al. (2020)
Parazoanthus swiftii	MISE JDR150614-118	St Eustatius, The Netherlands	MN873597	Poliseno et al. (2020)
Parazoanthus elongatus	MISE 170619	Chile	MN873596	Poliseno et al. (2020)
Hydrozoanthus gracilis	MISE JDR2016-hg1	Okinawa, Japan	MN873589	Poliseno et al. (2020)
Nanozoanthus harenaceus	MISE JDR201705Oura-nh1	Okinawa, Japan	MN873591	Poliseno et al. (2020)
Microzoanthus occultus	MISE JDR201705Oura-mo1	Okinawa, Japan	MN873590	Poliseno et al. (2020)
Neozoanthus aff. uchina	MISE JDR161218	Iriomote, Japan	MN873592	Poliseno et al. (2020)
Savalia savaglia	NA	Embiez Islands, France	DQ825686	Sinniger et al. (2007)
Stichopathes luetkeni	NA	NA	NC018377	Kayal et al. (2013)
Myriopathes japonica	NA	NA	NC027667	H.S. Kwak, E.H. Choi & U.W. Hwang, unpublished

1

Figure 1

Figure 1 . Bayesian inference phylogenetic tree of Zoantharia based on the concatenation of 13 mitochondrial protein-coding genes.

The phylogenetic trees computed with the Bayesian and the maximum-likelihood methods resulted in the same topologies, and hence node supports are displayed in posterior probabilities and bootstrap values.

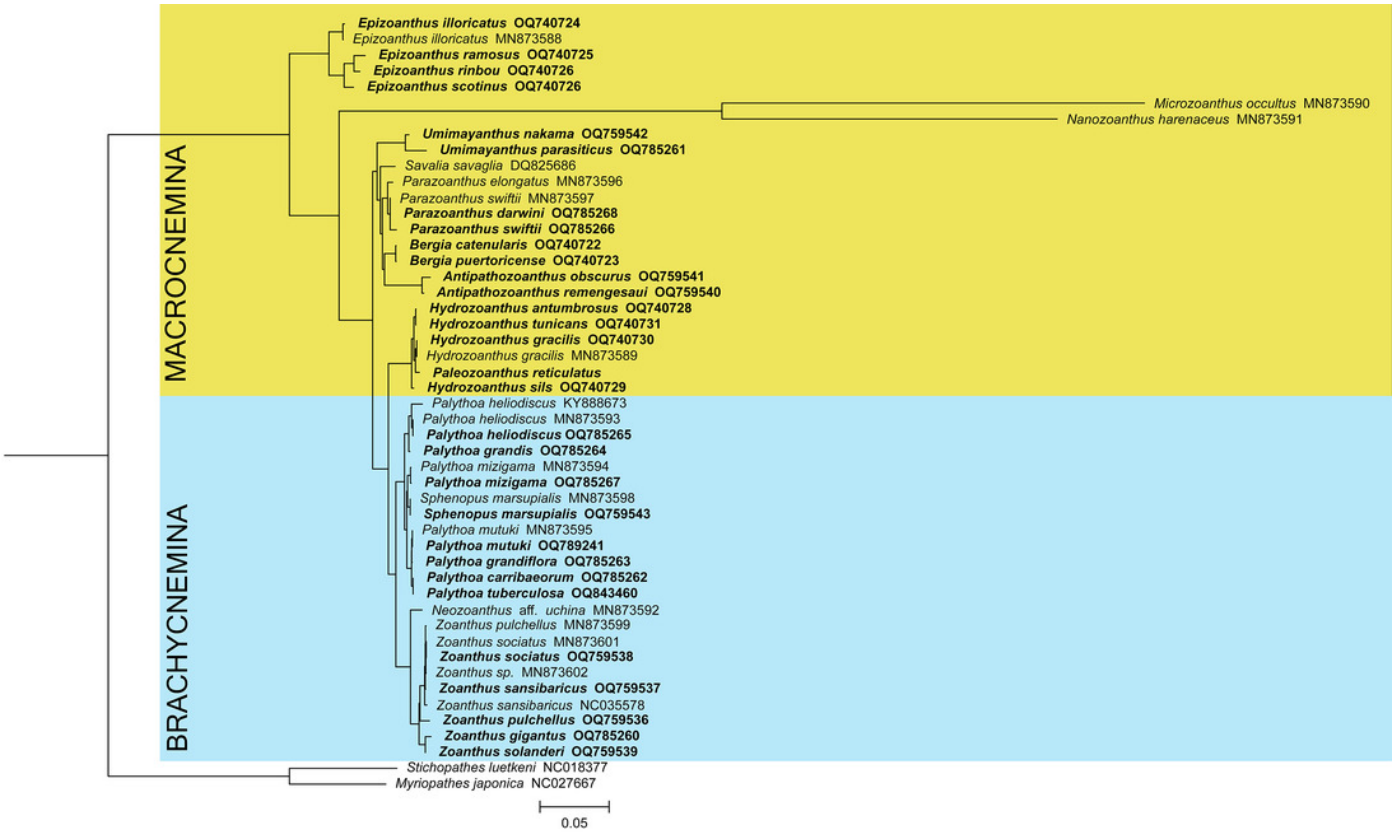


Figure 2

Figure 2 . Phylogenetic relationships of 18 zoantharian species with their repeat class abundance and respective genome size.

A) Cladogram of zoantharian phylogeny, B) repeat class abundance, C) genome size.

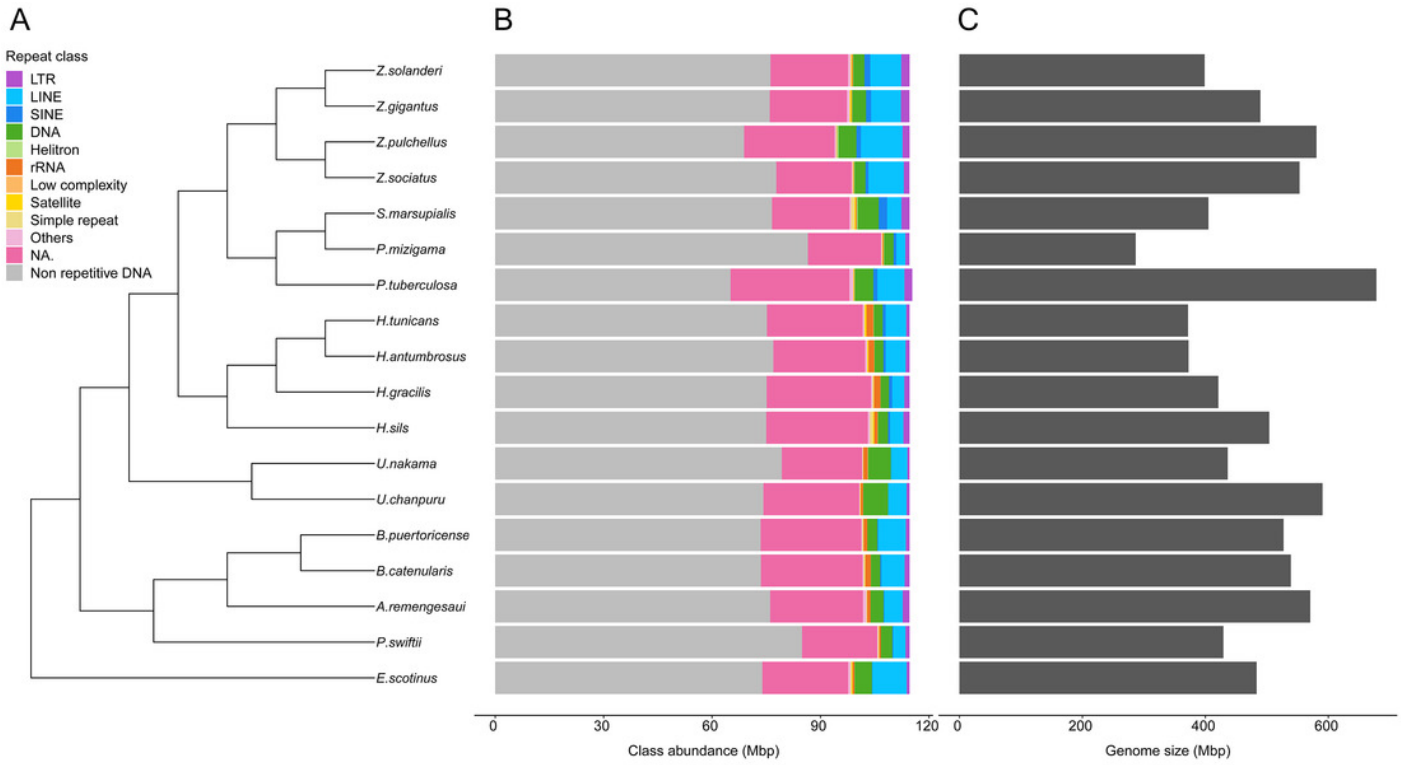


Figure 3

Figure 3. Heatmap representing transposable elements family abundance in 18 species of zoantharians.

TEs absent from a given species genomes are represented in cells with grey background.

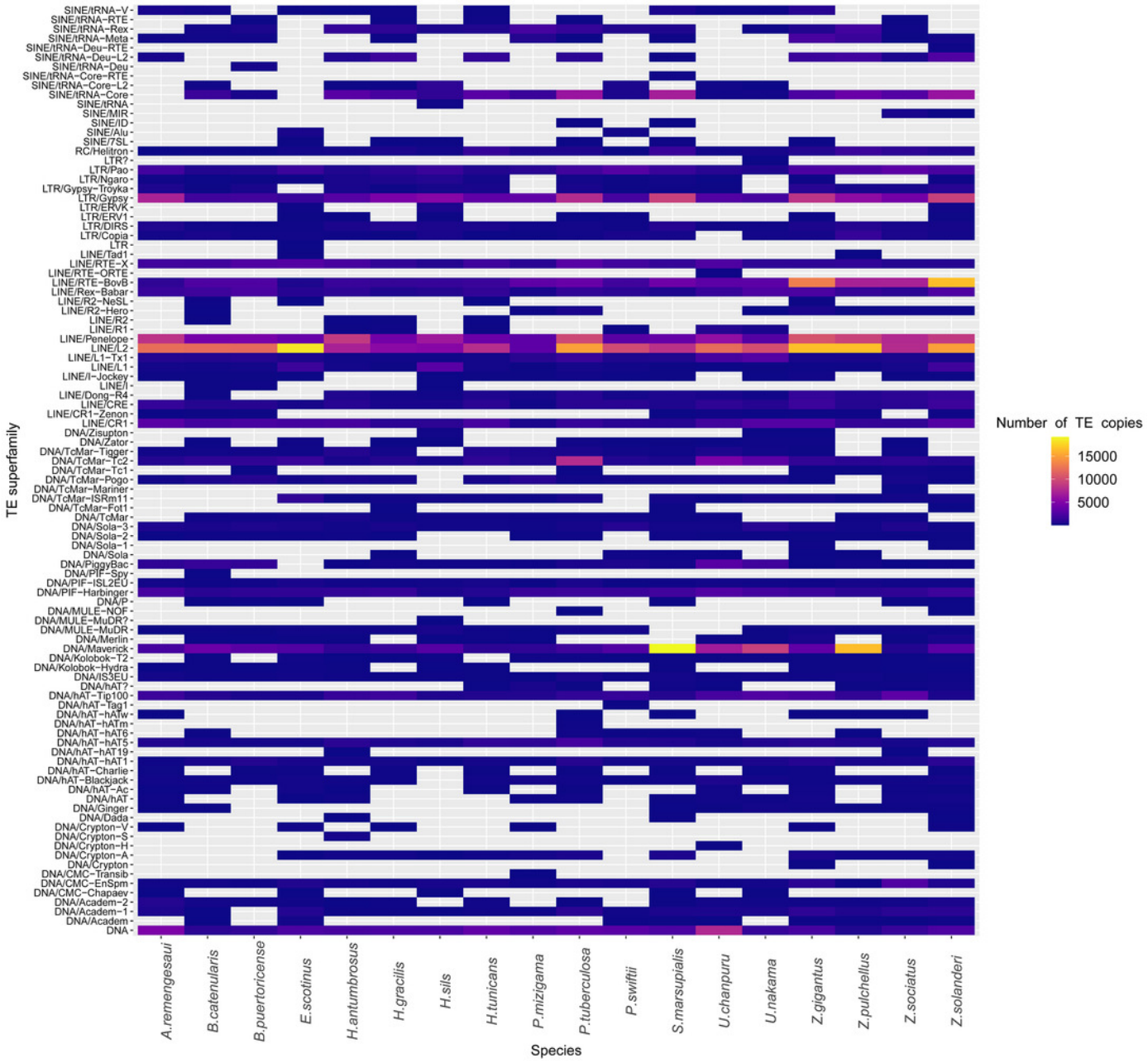


Figure 4

Figure 4. Transposable elements divergence landscapes for 18 species of zoantharians.

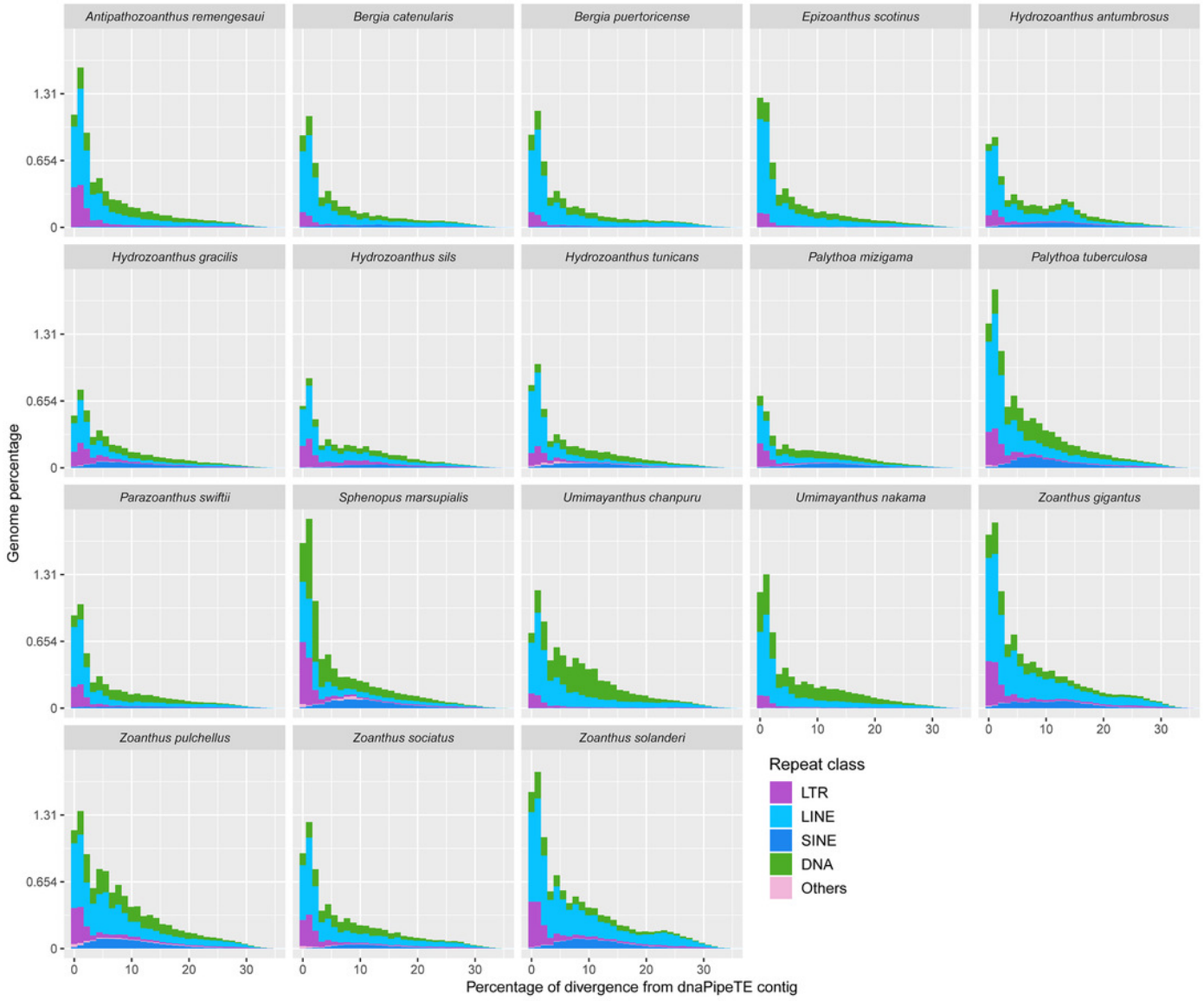


Figure 5

Figure 5. Cluster sizes and annotations normalized by genome sizes among repeated elements of 18 zoantharian species.

Species names are shown as three letter codes. *U*. *chanpuru*: UCH; *U*. *nakama*: UNA; *P*. *swiftii*: PSW; *P*. *tuberculosa*: PTU; *P*. *mizigama*: PMI; *A*. *remengesaui*: ARE; *Z*. *solanderi*: ZSL; Z. *gigantus*: ZGI; *H*. *tunicans*: HTU; *H*. *antumbrosus*: HAN; *H*. *gracilis*: HGR; *H*. *sils*: I; Z. *pulchellus*: ZPU; *B*. *catenularis*: BCA; *B*. *puertoricense*: BPU; *E*. *scotinus*: ESC; *S*. *marsupialis*: SMA; *Z*. *sociatus*: ZSC.

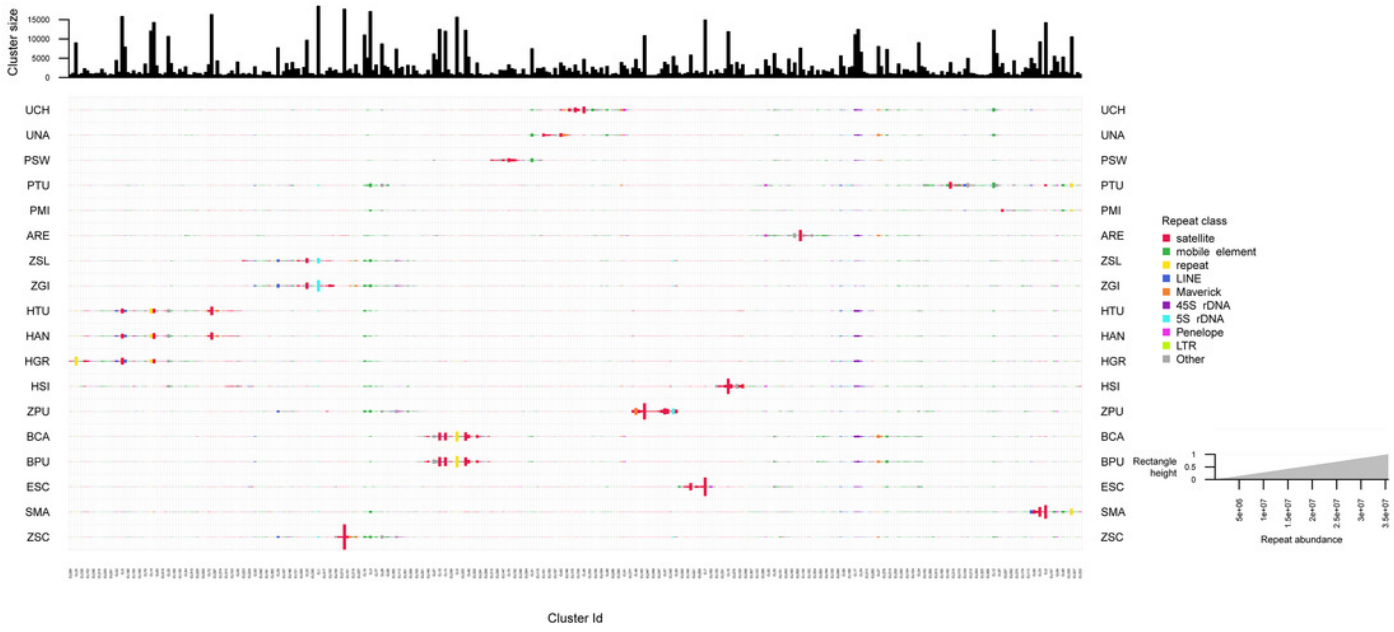
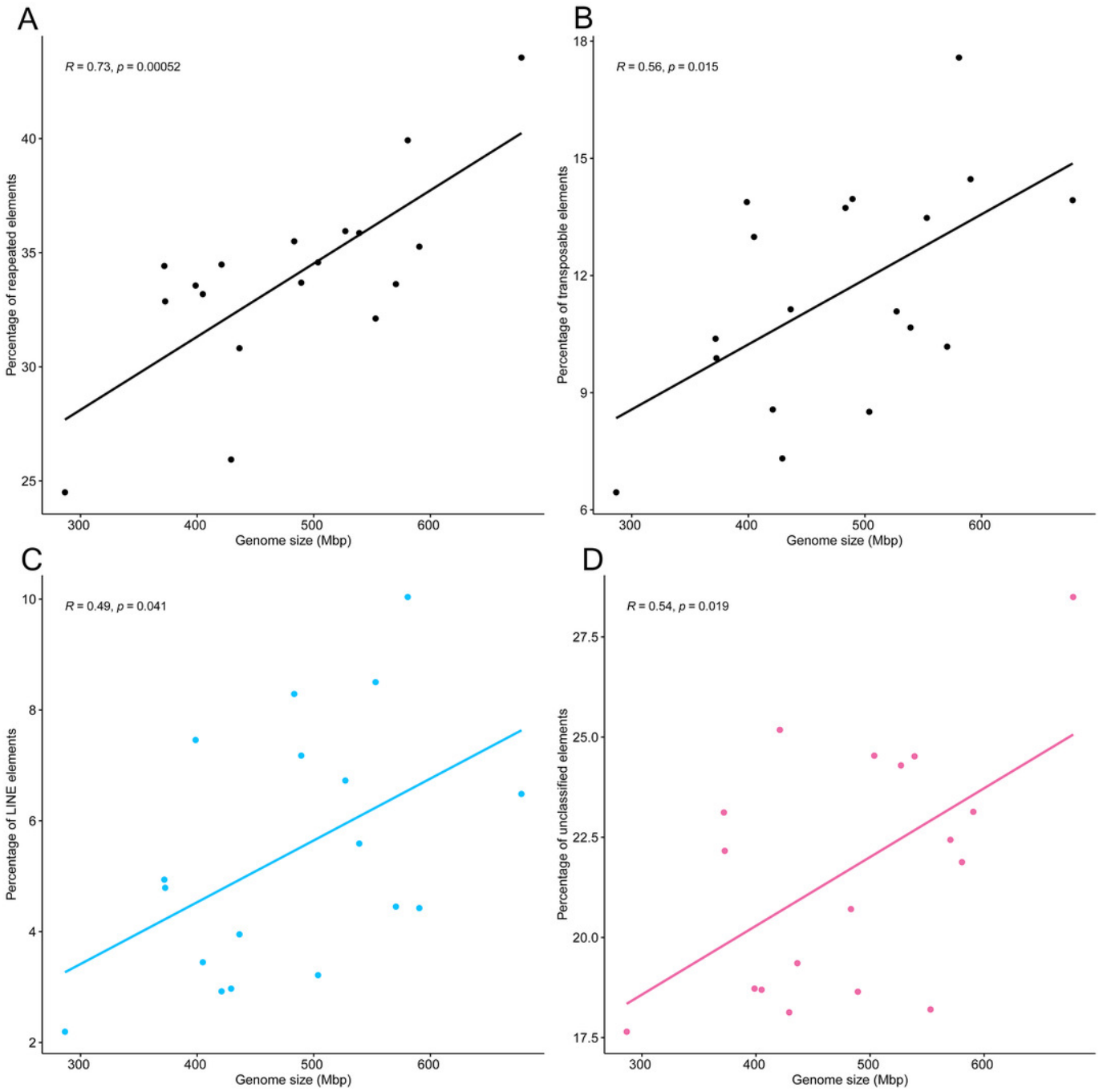


Figure 6

Figure 6. Pearson’s correlation between genome size of 18 zoantharian species and their respective percentage among categories of repeated DNA.

Pearson’s correlations between genome sizes and percentages of A) total repeated elements, B) transposable elements, C) LINEs elements, and D) unclassified repeats


A
R = 0.73, p = 0.00052
Percentage of reapeated elements
Genome size (Mbp)
B
R = 0.56, p = 0.015
Percentage of transposable elements
Genome size (Mbp)
C
R = 0.49, p = 0.041
Percentage of LINE elements
Genome size (Mbp)
D
R = 0.54, p = 0.019
Percentage of unclassified elements
Genome size (Mbp)