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 M. Maronna[1,5], Angelo Poliseno[1], Maria E.A. 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] 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 of America

[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 of America

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 2.4 fold magnitude. Significant correlations between genome size, repeated DNA content and transposable elements, respectively (Pearson's correlation test $R^2 = 0.47$, $p = 0.0016$; $R^2 = 0.22$, $p = 0.05$) 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 70%) could not be annotated 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.

INTRODUCTION

The C-value, the size of a species' haploid genome, has long been noticed to exhibit considerable variations that do not correlate with a given species expected complexity. This discrepancy has been referred to the "C-value paradox", highlighting the confusion regarding this pattern (*Thomas, 1971*; *Elliott & Gregory, 2015*). Indeed, despite genome sizes being in most cases relatively constant within species (*Swift, 1950*; *Gregory & Johnston, 2008*; *Dai et al., 2022*), intraspecific variation is well recognized (*Bonnivard et al., 2009*; *Stelzer, Pichler & Hatheuer, 2021a*) and large variations exist between closely related species (*Yuan et al., 2018*; *Wong et al., 2019*; *Becking, Gilbert & Cordaux, 2020*; *Shah, Hoffman & Schielzeth, 2020*; *Paule et al., 2021*). 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*). In particular, the relative content of transposable elements has been found to explain dynamics of interspecific genome sizes variations in many groups (*Kapusta, Suh & Feschotte, 2017*; *Lee & Kim, 2014*; *Wong et al., 2019*; *Becking, Gilbert & Cordaux, 2020*; *Lehmann et al., 2021*; *Meyer et al., 2021*). 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: type of elements involved, 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, several of which are potentially involved in genome size variations. 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) can move within a genome and are 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 long terminal repeats (LTRs), long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs). 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*). LTRs have been shown to largely impact genome sizes in plants (*Dai et al., 2022*) and salamanders (*Sun et al., 2012*),

LINEs have been involved in *Hydra* (*Wong et al., 2019*) and in the giant lungfish genome (*Meyer et al., 2021*), while the implicated group of elements were SINEs in larvaceans (*Naville et al., 2019*) and DNA transposons in fish (*Lehmann et al., 2021*).

Different effects on the genomes may be considered along with the different groups of transposable elements involved. 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, Pergner & Kozmik, 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. Considering the variety of elements involved in different taxa and their large array of potential implications, a better understanding of repeated elements and genome sizes requires further research on understudied groups (*Elliott & Gregory, 2015*). *Hotaling, Kelley & Frandsen (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. Many groups still lack basic genome size information, as seen in the genome size database (*Gregory, 2023*: https://www.genomesize.com/), where most groups of invertebrates have less than 100 recorded genome sizes, whereas fish, insects and mammals have several hundreds to greater than a thousand records. 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, Sinniger &*

Reimer, 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, Maeda & Reimer, 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*). Therefore, genomic data produced on zoantharians so far has mainly been employed to investigate phylogenetic relationships. However and in spite of high relevance of zoantharians in terms of evolution, ecology and biochemical potential, basic understanding of their genomes are lacking, including a lack of information about their genome sizes and repetitive element content.

To fill this gap we investigated genomic data of 32 species of zoantharians, spanning 10 genera of the order and five out of nine 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. 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 X-Ten, 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 five *Epizoanthus* species in the scope of the present paper have been already presented (*Kise, Reimer & Pirro, 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 an Illumina Hi-Seq X platform using a 150 pair-end chemistry.

The sequencing experimental data are available on the Sequence Read Archive (SRA) with accession numbers and corresponding information on the specimen collection are listed in Table S1. 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 v. 0.11.9 (*Andrews, 2010*). Prior to the following analyses, paired-end reads adapter sequences were removed in Trimmomatic v. 0.39 with default parameters (*Bolger, Lohse & Usadel, 2014*).

Mitochondrial genome assembly and phylogeny

Mitochondrial genomes (mtDNA) were assembled *de-novo* with NOVOPlasty v. 3.8.3 (*Dierckxsens, Mardulyn & Smits, 2017*), with a *k*-mer size comprised from 29 to 33. A partial *COI* sequence (~780 bp) 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 sequence of phylogenetically close species (Accession numbers of sequences used as seeds are listed in Table S2). Although the assembly was performed *de novo*, the input of a reference genome facilitates the process, hence 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*). Annotation was done using the Predict and Annotate tool by comparing mitogenomes with a reference mitogenome annotation of *Palythoa heliodiscus* (NC035579) 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. This same method was used to annotate tRNA and rRNA genes.

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 S3 along with two antipatharians mitogenome assemblies that were used as outgroups in the phylogenetic trees (*Kayal et al., 2018*, Kwak, Choi *et* Hwang, unpublished). The thirteen alignments were concatenated in Sequence Matrix v.1.8 (*Vaidya, Lohman & Meier, 2011*), resulting in 11,933 bp matrix. The best fitting evolutionary model of each gene was assessed with MEGA X (*Kumar, Stecher & Tamura, 2016*) using the AIC criterion (*Akaike, 1973*). Sequence evolution models were HKY+G+I (*Hasegawa, Kishino & Yano, 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, which was in both cases HKY+G+I.

Based on the concatenated alignment, phylogenetic trees were computed following the maximum-likelihood method (ML) 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 1,000 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 1,000 steps during $10 \ 10^6$ 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 ML 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 & Kingsford, 2011*) with the default k-mer size of 21. The resulting histograms were then input in GenomeScope (*Vurture et al., 2017*), 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., 2011*) and then annotates the resulting contigs with Repeatmasker (*Smit, Hubley & Green, 2013-2015*, RepeatMasker Open 4.0.7) and RepBase (*Bao, Kojima & Kohany, 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:140), as well as an average quality below 20, instead of the default 15 (SLIDNGWINDOW:4:40). 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 establish whether Symbiodiniaceae DNA contamination affected the results of genome sizes estimation and repeated elements analysis, contamination was checked using FastQScreen v.0.14.1 (*Wingett & Andrews, 2018*). For each species' dataset, approximately 100,000 reads (strictly trimmed, with mitochondrial genomes removed read datasets) were sampled and mapped against available genome assemblies of Symbiodiniaceae. The assemblies included one *Cladocopium* sp. and one *Symbiodinium* sp. Clade A3 generated by *Shoguchi et al. (2018)*, as well as *Breviolum minutum* (*Shoguchi et al., 2013*), *Durusdinium trenchii* (*Shoguchi et al., 2021*), *Symbiodinium microadriaticum* (*Aranda et al., 2016*), *Cladocopium proliferum* (*Liu et al., 2018*; genome presented as *C. goreaui* but species later named

C. proliferum in *Butler et al., 2023*), *Fugacium kawagutii* (*Liu et al., 2018*), and the human genome assembly GRCh37 (*Church et al., 2011*). FastQscreen outputs the number of reads from the dataset that mapped to reference genomes. When reads map several times over multiple genomes, this may indicate that instead of true contamination, very similar regions exist between some the reference assemblies and the target dataset. On the other hand, a large amount of reads mapping to a single genome must be interpreted as a serious case of contamination to consider (*Wingett & Andrews, 2018*).

To produce comparable estimates of repeated elements between species, the "fixed read sampling size" method was used (as opposed to using genome coverage). To determine the appropriate number of reads to sample, tests were run 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 S4). 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 S4). 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 764,453 for all species. In addition, the minimum size of contig to be included was set to 400 bp. 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.

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, Neumann & Macas, 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, Neumann & Macas, 2020*). The clustering of the reads was performed in comparative mode using the RepeatExplorer 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, Neumann & Macas (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, Neumann & Macas, 2020*), as this structure represents several different insertion sites of a transposon. Clusters that did not display such pattern remained unannotated. Finally, in order to visualize clusters that were shared or not shared between species, the corrected version of the cluster annotation file was input in RepeatExplorer visualizing tool.

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 (*Bianchini & Sánchez-Baracaldo, 2023*). 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, Sinniger & Reimer, 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*). Only the transposable elements classified as such were included into the analyses. Normality of residuals distribution was assessed with a Shapiro–Wilk test and the plots produced by the lm function were examined to ensure that datasets met the conditions required for the Pearson correlation test. Final plots were generated with ggscatter from the ggplot2 package (*Wickham, 2016*). Furthermore, phylogenetic independent contrasts were calculated using the function pic of the ape package (*Paradis & Schliep, 2019*) and similarly tested with Pearson correlation test, in order to assess whether correlations are maintained after correction for phylogeny. 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 *COI* 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 19,386 bp for *Epizoanthus rinbou* and 23,133 bp for *Umimayanthus parasiticus*. A
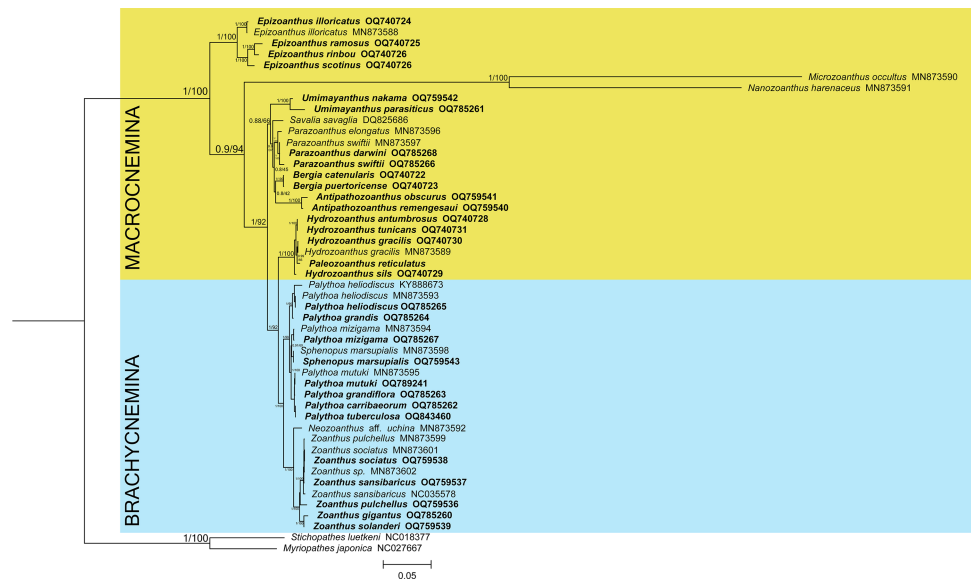
Figure 1 Bayesian inference phylogenetic tree of Zoantharia based on the concatenation of 13 mito-chondrial protein-coding genes. The phylogenetic trees computed with the Bayesian and the maximum-likelihood methods resulted in the same topologies, and hence node supports are displayed in posterior probabilities and bootstrap values. The two suborders in Zoantharia, Macrocnemina and Brachycnemina, are highlighted in yellow and light blue, respectively. Species names displayed in bold correspond to mito-chondrial genome data added by the present research.

Full-size ⬚ DOI: 10.7717/peerj.16188/fig-1

table summarizing the sizes of all complete mtgenomes is available in the supplementary material (Table S2).

The phylogenetic reconstructions performed with Bayesian inference and maximum-likelihood methods (Fig. 1) found the suborder Brachycnemina (highlighted in light blue) to be monophyletic with high support (Bayesian posterior probabilities = 1, maximum-likelihood bootstrap = 100%). Conversely, Macrocnemina (highlighted in yellow) was retrieved as paraphyletic, containing Brachycnemina as a sister clade of the macrocnemic genus *Hydrozoanthus*. 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 (Table S5). 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. The additional 12 species for which genome sizes could not be estimated also showed profiles where GenomeScope "model did not
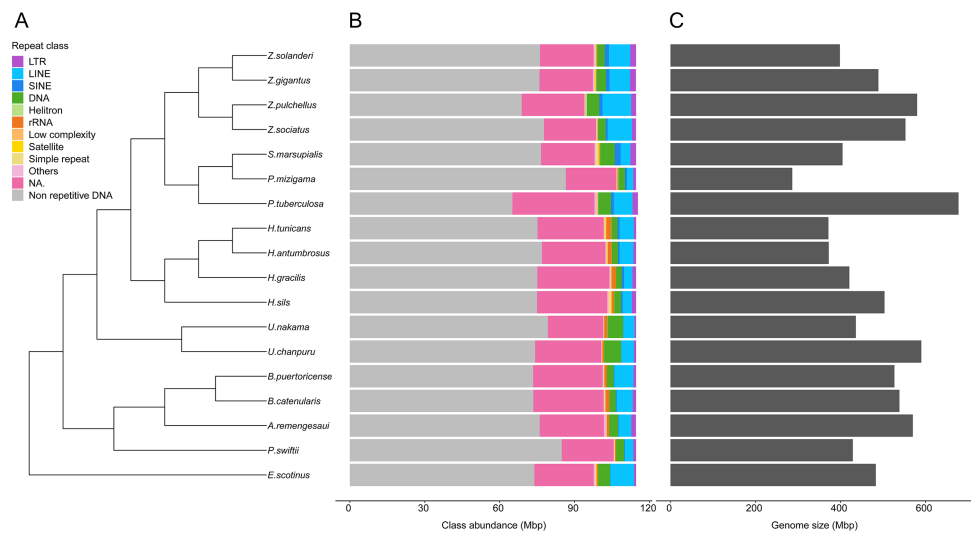
Figure 2 **Phylogenetic relationships of 18 zoantharian species with their repeat class abundance and respective genome size.** (A) Cladogram of zoantharian phylogeny, (B) repeat class abundance, (C) genome size. "NA" refers to non-annotated repeated elements.

converge", pointing to the absence of a k-mer peak in the data. Genome size of zoantharians species ranged between 286 and 678 million base pairs (Mbp) (Fig. 2). The genera *Zoanthus*, *Umimayanthus* and *Hydrozoanthus* overlapped in range with genome sizes between 370 Mbp and 590 Mbp, and maximum genome size differences within genus of 160 Mbp. Genus *Palythoa*, however, comprised the maximum size differences at the scale of the order with a 2.4 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 (Figs. 2, S1). The read abundance for each repeated element class and species are reported in Table S6. Despite similar total repeat abundances, the proportions of repeat classes seemed to vary (Figs. 2, S1). Of all repeated elements, up to 30 Mbp (~70% of 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 (Figs. 2, 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 S7). LINE/RTE-BovB 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
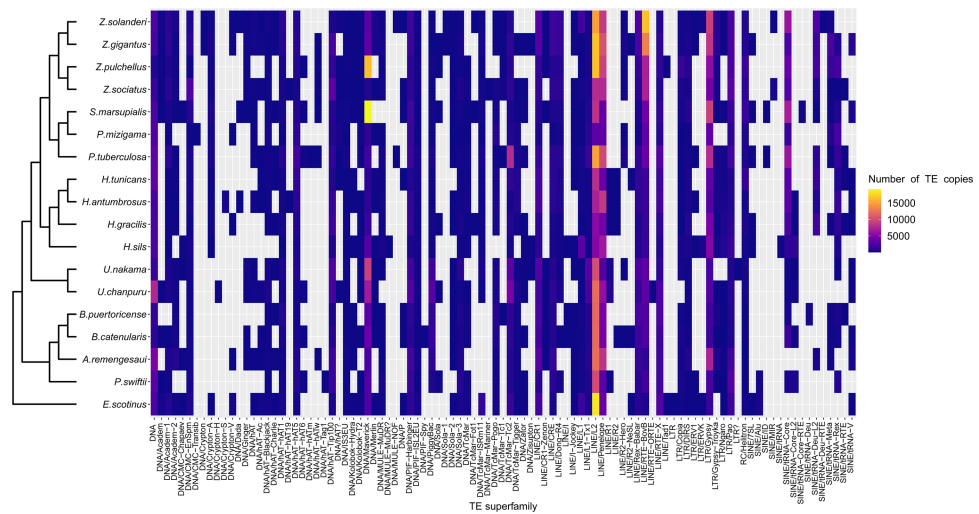
Figure 3 **Heatmap representing transposable elements family abundance in 18 species of zoantharians.** TEs absent from a given species genomes are represented in cells with grey background.

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 S8).

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 at a high percentage of divergence. The second spike was stronger in *H. antumbrosus* which displayed an increased abundance of LINEs 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 at this degree of divergence. *Zoanthus solanderi* also displayed a small bump related to the activity of LINEs 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 LINEs 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, LTRs appeared to have higher abundances, whereas SINEs disappeared, being at their peak abundance
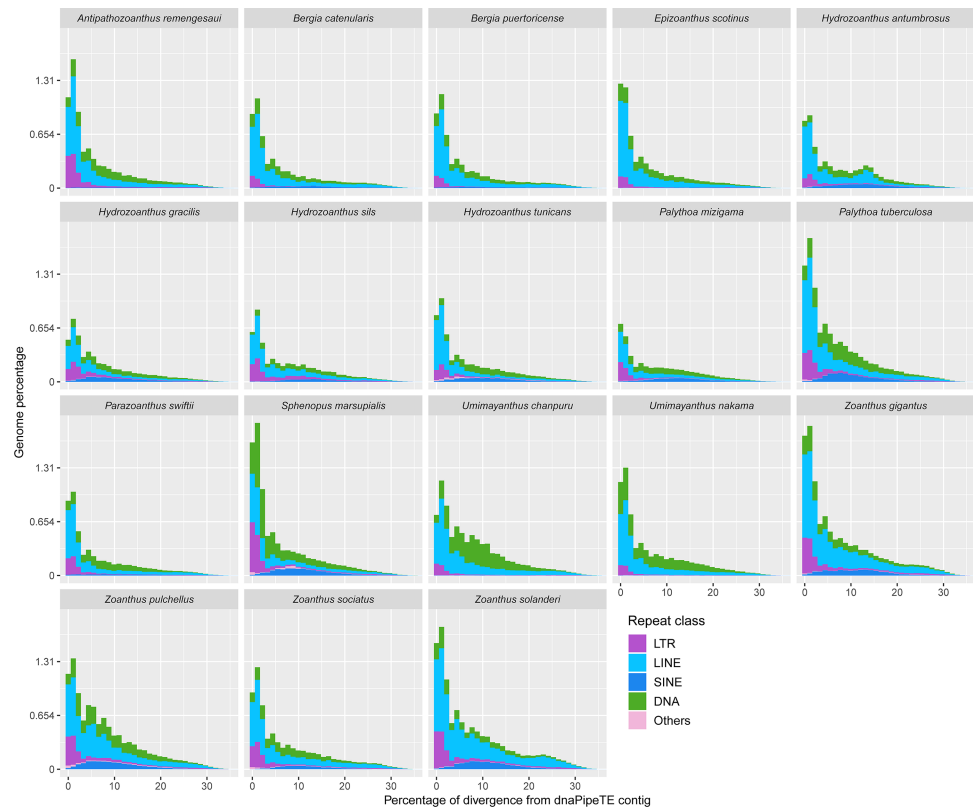
Figure 4   Transposable elements divergence landscapes for 18 species of zoantharians.
Full-size ☐ DOI: 10.7717/peerj.16188/fig-4

(~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.

The repeated elements clustering in RepeatExplorer2 resulted in the analysis of 4,929,668 reads, of which ~60% were assigned to 340 superclusters, and 354 clusters. Total number of reads detected in each repeat class are summarized in Table S9. Many clusters were represented by all zoantharian species, in particular clusters displayed in Figs. 5 and 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 (Figs. 5, 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, as shown by their small repeat abundance on Fig. 5 (*e.g.*, clusters 13, 34, 343). 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 (clusters 5 to 56). These groups of clusters corresponded essentially to satellite elements in
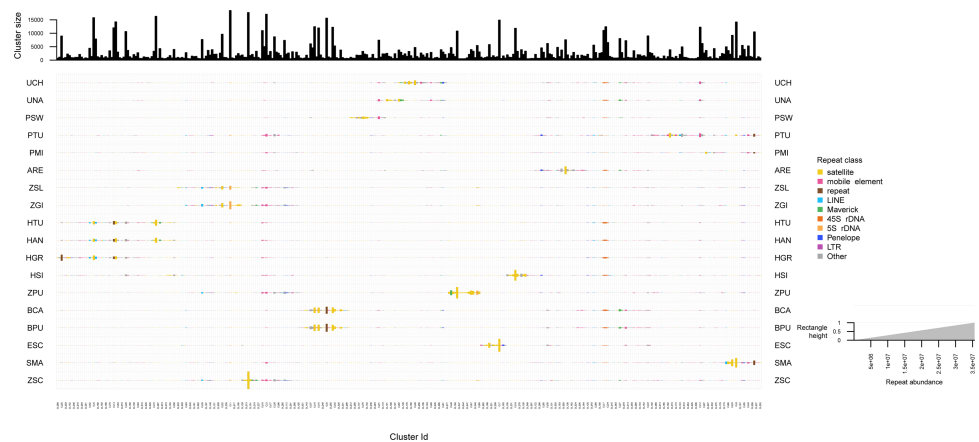
Figure 5 **Cluster sizes and annotations normalized by genome sizes among repeated elements of 18 zoantharian species.** Species names are shown as three letter codes. *U. chanpuru*: UCH; *U. nakama*: UNA; *P. swiftii*: PSW; *P. tuberculosa*: PTU; *P. mizigama*: PMI; *A. remengesaui*: ARE; *Z. solanderi*: 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.

Full-size 🖾 DOI: 10.7717/peerj.16188/fig-5

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*, and *S. marsupialis*) (Figs. 2 and 5). *Z. pulchellus* and *Z. sociatus* had the highest genome sizes in *Zoanthus* (580 and 553 Mb 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).

The screening of symbiont DNA in zoantharians datasets are shown in Fig. S4. Although overall, approximately 5% of reads mapped with symbiont genomes, for most species less than 1% of them mapped uniquely to a symbiont genome. In addition, the reads from several species associated with zooxanthellae did not map to Symbiodiniaceae genomes more than azooxanthellate zoantharians, suggesting low levels of contamination. The datasets where relatively high contamination levels were detected include *Zoanthus sansibaricus* (∼9.43% of reads mapping singularly to *Cladocopium proliferum*), *Palythoa carribaeorum* (4.3%), *Palythoa grandis* (3.4%) and *Palythoa grandiflora* (4.3%), which were not included in main analyses.

Correlation tests between genome size and repeated elements

Pearson's correlation test showed a correlation between the 18 genome sizes and the proportions of repeated elements, supported by an R^2 of 0.47 and a highly significant *p*-value of 0.0016. A weaker but statistically significant correlation was found between genome sizes and the percentage of transposable elements ($R^2 = 0.22$, $p = 0.05$), in which points appeared more dispersed (Figs. 6A, 6B). The tests of phylogenetic independent
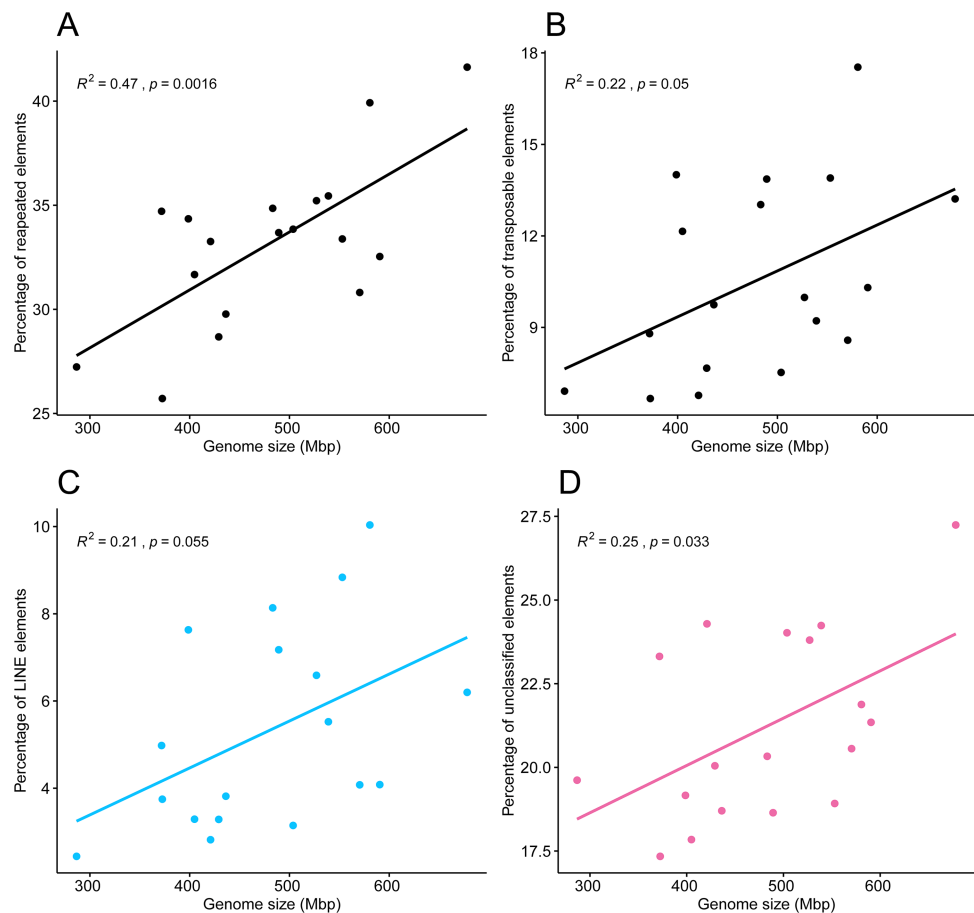
Figure 6 **Pearson's correlation between genome size of 18 zoantharian species and their respective percentage among categories of repeated DNA.** Pearson's correlations between genome sizes and percentages of (A) total repeated elements, (B) transposable elements, (C) LINEs elements, and (D) unclassified repeats.

Full-size 🖼 DOI: 10.7717/peerj.16188/fig-6

contrasts of genome sizes revealed significant relationships with the ones of same repeated element categories; repeated elements ($R^2 = 0.73$, $p = 1.3e{-}05$), transposable elements ($R^2 = 0.37$, $p = 0.0096$), LINEs ($R^2 = 0.28$, $p = 0.028$), and unclassified elements ($R^2 = 0.43$, $p = 0.0044$). 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.

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, Chevaldonné & Pawlowski (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 S1), 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 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, Sinniger & Reimer, 2016*). Interestingly, the present phylogenetic reconstruction placed *Paleozoanthus reticulatus* within genus *Hydrozoanthus* (Fig. 1), which belongs to the same family as *Terrazoanthus* (*Kise, Maeda & Reimer, 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*; *Kise et al, 2023*). 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, Sinniger & Reimer, 2015*). However, samples from another 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 estimates 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* (MD 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 correlated to the respective percentages of repeated and transposable elements (Figs. 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., 2021*). 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, Masucci & Reimer, 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 70% 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 may have contributed to the large amounts of unclassified repeats. However, annotation is likely the main explanation, as our assemblies' N50 and contig numbers (Table S10) were comparable to or better than those presented by the developers of dnaPipeTE (*Goubert et al., 2015*), who obtained significantly fewer unclassified elements.

Although we obtained large proportions of unclassified repeats in the dnaPipeTE analyses, the clustering and repeat annotation performed *via* RepeatExplorer2 suggested that they may be partly represented by satellite elements (Figs. 5, S3). Indeed, they accounted for ∼30% of the annotated elements in the comparative analysis (Table S9), yet they were almost absent from annotations *via* dnaPipeTE (Fig. 2). Conversely, numerous mobile elements could not be annotated from RepeatExplorer2. While this partly reflects the different sensitivities of the two pipelines and the databases that they use, the consistently high amounts of unclassified repeats in zoantharians highlight that much remains to be discovered with regards to their genomes. More efforts into assembling and characterizing their repeatomes will surely reveal interesting elements. Indeed, the percentages of unclassified repeat categories were found to be correlated to genome size (Fig. 6D), suggesting that elements with significance for genome size dynamics are contained among unclassified repeats.

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

The literature on the roles of repeated elements in genome sizes has largely focused on cases displaying extreme genome size variations. In these situations, dramatic changes of genome sizes in association with a single specific repeated class have been reported. Notably, the class of repeated elements involved varies between taxa; in larvaceans SINEs appear to drive genome size increases (*Naville et al., 2019*), while satellites and Helitrons were the main contributors in migratory locusts (*Shah, Hoffman & Schielzeth, 2020*). In *Hydra,* LINEs have had a major expansion event leading to dramatic genome size increase in the subgroup of brown *Hydra* (*Wong et al., 2019*). Although different repeated elements are clearly involved in genome size dynamics in different groups, the degree of variation between taxa is not well understood. The present dataset offers insights in this question

by adding to the knowledge of repeated elements in Cnidaria. Genome size variations in zoantharians do not appear to be as important as in *Hydra*, but still reach a maximum variation of 2.4 fold, between congeners *Palythoa tuberculosa* and *P. mizigama*. However, it is notable that LINEs–the class responsible for genome size expansion in *Hydra*– were consistently one of the most abundant in our dataset (Fig. 2B), and that a significant correlation between this class and genome size was detected (Fig. 6C). Furthermore, the repeat landscapes of most species showed a high number of LINEs with low divergence (Fig. 4). Such patterns have been interpreted as a sign of recent TE activity; TE copies in the 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 (Figs. 4, 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 may have led to increased genome sizes in zoantharians, and potentially may have done so across Cnidaria. However, LINEs were found in varying, yet relatively small abundance overall and did not account for all the observed genome size variations on their own. Transposable elements are known to move other regions with them (*Langer et al., 2007*; *Qiu & Köhler, 2020*), and LINEs may have been transporting other regions of the genome with them, regions that could similarly account for the genome size discrepancies. In parallel with their effects on genome size, LINE/L2 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 (Figs. 5, 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, Hoffman & Schielzeth, 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, Hoffman & Schielzeth, 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. Considering our results on LINEs, unclassified and satellite elements, 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, 2020*). In the first theory, the accumulation of DNA is considered 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., 2021b*), 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 influence 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, Allemand & Weis, 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 2.4 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* results, 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, Sinniger & Reimer, 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 elements of their symbiotic counterpart. Although there is no documented evidence of TE transfer between Symbiodiniacae and hosts, transposable elements transcripts in *Symbiodinium* have been shown to be upregulated in situations of environmental stress (*Chen et al., 2018*). In addition, symbiotic *Symbiodinium* species have been found to have a larger content of transposable elements than free-living counterparts (*González-Pech et al., 2021*). Considering the abundance and activity of TEs in Symbiodiniacae, horizontal transfer may have contributed to the large genome size observed in the case of *P. tuberculosa*.

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

CONCLUSIONS

In this article, we explored the relationships between phylogeny, genome size variations, and the repetitive elements composition of a scarcely studied group of cnidarians. Our results show that genome sizes observed in zoantharians are likely the product of complex historical dynamics of the repeatome. We found a high number of unknown repeats with potential implications in genome size. Recent expansion events of LINEs, DNA and satellite elements were identified in multiple species, raising questions on the role of these elements in genome evolution of cnidarians and the consequences of their activity. Until now no information was available for zoantharian genome sizes, and we here present such information for 18 specimens from five of the nine zoantharian families. This research demonstrates the power of next-generation sequencing projects aimed at understudied taxa, allowing a rapid increase in our basic understanding of such poorly studied groups. This

sequencing project also allowed us to clarify the phylogenetic position of *Paleozoanthus via* analyses of an old specimen; such work could very likely not have been performed utilizing traditional genetic methods. Finally, as there are notable questions related to the ecology, symbioses, development and evolution of zoantharians, the genome data and repeatome characteristics presented here will serve as important baseline data to investigate such questions in future genomic projects.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

James D. Reimer is an Academic Editor for PeerJ. Stacy Pirro is employed by Iridian Genomes.

Author Contributions

- Chloé Julie Loïs Fourreau conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Hiroki Kise performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Mylena Daiana Santander conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Stacy Pirro performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Maximiliano M. Maronna conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Angelo Poliseno conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Maria E.A. Santos analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- James Davis Reimer conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The SRA and Genbank accession numbers for each presented sequencing data and mitochondrial genome are available in Genbank:

Antipathozoanthus remengesaui, PRJNA598175; *Antipathozoantus obscurus*, PRJNA598176; *Bergia catenularis*, PRJNA662740; *Bergia puertoricense*, PRJNA662980; *Epizoanthus illoricatus*, PRJNA598181; *Epizoanthus planus*, PRJNA676135; *Epizoanthus ramosus*, PRJNA598173; *Epizoanthus rinbou*, PRJNA662986; *Epizoanthus scotinus*, PRJNA662773; *Hydrozanthus tunicans*, PRJNA645597; *Hydrozoanthus antumbrosus*, PRJNA662983; *Hydrozoanthus gracilis*, PRJNA662988; *Hydrozoanthus sils*, PRJNA662735; *Paleozoanthus reticulatus*, PRJNA622546; *Palythoa carribaeorum*, PRJNA598184; *Palythoa grandiflora*, PRJNA598185; *Palythoa grandis*, PRJNA580275; *Palythoa heliodiscus*, PRJNA598194; *Palythoa mizigama*, PRJNA957836; *Palythoa mutuki*, PRJNA598187; *Palythoa tuberculosa*, PRJNA946699; *Parazoanthus swiftii*, PRJNA662982; *Parazoantus darwini*, PRJNA662981; *Sphenopus marsupialis*, PRJNA662993; *Umimayanthus chanpuru*, PRJNA662702; *Umimayanthus nakama*, PRJNA645598; *Umimayanthus parasiticus*, PRJNA662764; *Zoanthus gigantus*, PRJNA598193; *Zoanthus pulchellus*, PRJNA645596; *Zoanthus sociatus*, PRJNA598186; *Zoanthus solanderi*, PRJNA662769; *Zoanthus sansibaricus*, PRJNA598188.

Data Availability

The following information was supplied regarding data availability:

The raw data is available in the Supplementary Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.16188#supplemental-information.

REFERENCES

Adachi K, Miyake H, Kuramochi T, Mizusawa K, Okumura S. 2017. Genome size distribution in phylum Cnidaria. *Fisheries Science* **83(1)**:107–112 DOI 10.1007/s12562-016-1050-4.

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

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215(3)**:403–410 DOI 10.1016/S0022-2836(05)80360-2.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. *Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.*

Aranda M, Li Y, Liew YJ, Baumgarten S, Simakov O, Wilson MC, Ashoor H, Bougouffa S, Bajic VB, Ryu T, Ravasi T, Bayer T, Micklem G, Kim H, Bhak J, Lajeunesse TC, Voolstra CR. 2016. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports* **6**:39734 DOI 10.1038/srep39734.

Aratake S, Taira Y, Fujii T, Roy MC, Reimer JD, 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 DOI 10.1016/j.toxicon.2015.12.004.

Baack EJ, Whitney KD, Rieseberg LH. 2005. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *New Phytologist* **167(2)**:623–630 DOI 10.1111/j.1469-8137.2005.01433.x.

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6(1)**:11 DOI 10.1186/s13100-015-0041-9.

Barrett NJ, Hogan RI, Allcock AL, Molodtsova T, Hopkins K, Wheeler AJ, 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 DOI 10.3389/fmars.2020.00440.

Becking T, Gilbert C, Cordaux R. 2020. Impact of transposable elements on genome size variation between two closely related crustacean species. *Analytical Biochemistry* **600**:113770 DOI 10.1016/j.ab.2020.113770.

Bianchini G, Sánchez-Baracaldo P. 2023. TreeViewer - Cross-platform software to draw phylogenetic trees (Version 2.1.0). *Available at https://treeviewer.org/.*

Blommaert J. 2020. Genome size evolution: towards new model systems for old questions. *Proceedings of the Royal Society B: Biological Sciences* **2871933**:20201441 DOI 10.1098/rspb.2020.1441.

Blommaert J, Riss S, Hecox-Lea B, Mark Welch DB, Stelzer CP. 2019. Small, but surprisingly repetitive genomes: transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species complex. *BMC Genomics* **20**:466 DOI 10.1186/s12864-019-5859-y.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30(15)**:2114–2120 DOI 10.1093/bioinformatics/btu170.

Bonnivard E, Catrice O, Ravaux J, Brown SC, Higuet D. 2009. Survey of genome size in 28 hydrothermal vent species covering 10 families. *Genome* **52(6)**:524–536 DOI 10.1139/G09-027.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL, Feschotte C. 2018. Ten things you should know about transposable elements. *Genome Biology* **19(1)**:199 DOI 10.1186/s13059-018-1577-z.

Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner (LBNL-7065E). Berkeley: Lawrence Berkeley National Laboratory (LBNL).

Butler CC, Turnham KE, Lewis AM, Nitschke MR, Warner ME, Kemp DW, Hoegh-Guldberg O, Fitt WK, van Oppen MJH, LaJeunesse TC. 2023. Formal recognition of host-generalist species of dinoflagellate (*Cladocopium*, Symbiodiniaceae) mutualistic with Indo-Pacific reef corals. *Journal of Phycology* **59**:698–711 DOI 10.1111/jpy.13340.

Chen J, Cui G, Wang X, Liew YJ, Aranda M. 2018. Recent expansion of heat-activated retrotransposons in the coral symbiont *Symbiodinium microadriaticum*. *The ISME Journal* **12**:639–643 DOI 10.1038/ismej.2017.179.

Chi SI, Johansen SD. 2017. Zoantharian mitochondrial genomes contain unique complex group I introns and highly conserved intergenic regions. *Gene* **628**:24–31 DOI 10.1016/j.gene.2017.07.023.

Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, Chen HC, Agarwala R, McLaren WM, Ritchie GR, Albracht D, Kremitzki M, Rock S, Kotkiewicz H, Kremitzki C, Wollam A, Trani L, Fulton L, Fulton R, Matthews L, Whitehead S, Chow W, Torrance J, Dunn M, Harden G, Threadgold G, Wood J, Collins J, Heath P, Griffiths G, Pelan S, Grafham D, Eichler EE, Weinstock G, Mardis ER, Wilson RK, Howe K, Flicek P, Hubbard T. 2011. Modernizing reference genome assemblies. *PLOS Biology* **9(7)**:e1001091 DOI 10.1371/journal.pbio.1001091.

Dai S, Zhu X, Hutang G, Li J, Tian J, Jiang X, Zhang D, Gao L. 2022. Genome size variation and evolution driven by transposable elements in the genus *Oryza*. *Frontiers in Plant Science* **13**:921937 DOI 10.3389/fpls.2022.921937.

Davy SK, Allemand D, Weis VM. 2012. Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiology and Molecular Biology Reviews* **76(2)**:229–261 DOI 10.1128/MMBR.05014-11.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* **45(4)**:e18 DOI 10.1093/nar/gkw955.

Dion-Côté AM, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Molecular Biology and Evolution* **31(5)**:1188–1199 DOI 10.1093/molbev/msu069.

Dudoit A, Santos MEA, Reimer JD, Toonen RJ. 2021. Phylogenomics of *Palythoa* (Hexacorallia: Zoantharia): probing species boundaries in a globally distributed genus. *Coral Reefs* **41(3)**:655–672 DOI 10.1007/s00338-021-02128-4.

Edgar C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32(5)**:1792–1797 DOI 10.1093/nar/gkh340.

Elliott TA, Gregory TR. 2015. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **3701678**:20140331 DOI 10.1098/rstb.2014.0331.

González-Pech RA, Stephens TG, Chen Y, Mohamed AR, Cheng Y, Shah S, Dougan KE, Fortuin MDA, Lagorce R, Burt DW, Bhattacharya D, Ragan MA, Xin Chan C. 2021. Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus *Symbiodinium*. *BMC Biology* **19(73)** DOI 10.1186/s12915-021-00994-6.

Goubert C, Modolo L, Vieira C, Valiente Moro C, Mavingui P, Boulesteix M. 2015. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biology and Evolution* **7(4)**:1192–1205 DOI 10.1186/s12915-021-00994-6.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* **29(7)**:644–652 DOI 10.1038/nbt.1883.

Gregory TR. 2005. The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Annals of Botany* **95(1)**:133–146 DOI 10.1093/aob/mci009.

Gregory TR. 2023. Animal genome size database. *Available at* http://www.genomesize.com.

Gregory TR, Johnston JS. 2008. Genome size diversity in the family Drosophilidae. *Heredity* **101**:228–238 DOI 10.1038/hdy.2008.49.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22(2)**:160–174 DOI 10.1007/BF02101694.

Hénault M, Souhir M, Charron G, Landry CR. 2020. The effect of hybridization on transposable element accumulation in an undomesticated fungal species. *eLife* **9**:e60474 DOI 10.7554/eLife.60474.

Hirose M, Obuchi M, Hirose E, Reimer JD. 2011. Timing of spawning and early development of *Palythoa tuberculosa* (Anthozoa, Zoantharia, Sphenopidae) in Okinawa, Japan. *The Biological Bulletin* **220(1)**:23–31 DOI 10.1086/BBLv220n1p23.

Hotaling S, Kelley JL, Frandsen PB. 2021. Toward a genome sequence for every animal: Where are we now? *Proceedings of the National Academy of Sciences of the United States of America* **118(52)**:e2109019118 DOI 10.1073/pnas.2109019118.

Irei Y, Sinniger F, Reimer JD. 2015. Descriptions of two azooxanthellate *Palythoa* species (Subclass Hexacorallia, Order Zoantharia) from the Ryukyu Archipelago, southern Japan. *ZooKeys* **478**:1–26 DOI 10.3897/zookeys.478.8512.

Ishikawa M, Yuyama I, Shimizu H, Nozawa M, Ikeo K, Gojobori T. 2016. Different endosymbiotic interactions in two hydra species reflect the evolutionary history of endosymbiosis. *Genome Biology and Evolution* **8(7)**:2155–2163 DOI 10.1093/gbe/evw142.

Jeffery NW, Hultgren K, Chak STC, Gregory TR, Rubenstein DR. 2016. Patterns of genome size variation in snapping shrimp. *Genome* **59(6)**:393–402 DOI 10.1139/gen-2015-0206.

Johansen SD, Chi SI, Dubin A, Jørgensen TE. 2021. The mitochondrial genome of the sea anemone *Stichodactyla haddoni* reveals catalytic introns, insertion-like element, and unexpected phylogeny. *Life* **11(5)**:402 DOI 10.3390/life11050402.

Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences of the United States of America* **114(8)**:E1460–E1469 DOI 10.1073/pnas.1616702114.

Kayal E, Bentlage B, Pankey SM, Ohdera AH, Medina M, Plachetzki DC, Ryan JF, Collins AG. 2018. Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *BMC Evolutionary Biology* **18**:68 DOI 10.1186/s12862-018-1142-0.

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(12)**:1647–1649 DOI 10.1093/bioinformatics/bts199.

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 DOI 10.1038/s41559-019-0853-y.

Kise H, Maeda T, Reimer JD. 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 DOI 10.1016/j.ympev.2018.10.011.

Kise H, Moritaki T, Iguchi A, Reimer JD. 2022. Epizoanthidae (Hexacorallia: Zoantharia) associated with *Granulifusus* gastropods (Neogastropoda: Fasciolariidae) from the Indo-West Pacific. *Organisms Diversity and Evolution* **22(3)**:543–554 DOI 10.1007/s13127-022-00552-0.

Kise H, Reimer JD, Pirro S. 2023. The complete genome sequences of 7 species of *Epizoanthus* (Epizoanthidae, Zoantharia, Hexacorallia, Cnidaria). *Biodiversity Genomes* **2023**:10.56179/001c.72780 DOI 10.56179/001c.72780.

Kise H, Santos MEA, Fourreau CJL, Iguchi A, Goto R, Reimer JD. 2023. Evolutionary patterns of host switching, lifestyle mode, and the diversification history in symbiotic zoantharians. *Molecular Phylogenetics and Evolution* **182**:107732 DOI 10.1016/j.ympev.2023.107732.

Kozlov AM, 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 DOI 10.1093/bioinformatics/btz305.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33(7)**:1870–1874 DOI 10.1093/molbev/msw054.

Langer M, Sniderhan LF, Grossniklaus U, Ray A. 2007. Transposon excision from an atypical site: a mechanism of evolution of novel transposable elements. *PLOS ONE* **2(10)**:e965 DOI 10.1371/journal.pone.0000965.

Lee S-I, Kim N-S. 2014. Transposable elements and genome size variations in plants. *Genomics & Informatics* **12(3)**:87–97 DOI 10.5808/GI.2014.12.3.87.

Lefébure T, Morvan C, Malard F, François C, Konecny-Dupré L, Guéguen L, Weiss-Gayet M, Seguin-Orlando A, Ermini L, Sarkissian CDer, Charrier NP, Eme D, Mermillod-Blondin F, Duret L, Vieira C, Orlando L, Douady CJ. 2017. Less effective selection leads to larger genomes. *Genome Research* **27(6)**:1016–1028 DOI 10.1101/gr.212589.116.

Lehmann R, Kovařík A, Ocalewicz K, Kirtiklis K, Zuccolo A, Tegner JN, Wanzenböck J, Bernatchez L, Lamatsch DK, Symonováand R. 2021. DNA transposon expansion is associated with genome size increase in mudminnows. *Genome Biology and Evolution* **3(10)**:evab228 DOI 10.1093/gbe/evab228.

Leinaas HP, Jalal M, Gabrielsen TM, Hessen DO. 2016. Inter- and intraspecific variation in body- and genome size in calanoid copepods from temperate and arctic waters. *Ecology and Evolution* **6(16)**:5585–5595 DOI 10.1002/ece3.2302.

Lin M-F, Kitahara MV, Luo H, Tracey D, Geller J, Fukami H, Miller DJ, Chen CA. 2014. Mitochondrial genome rearrangements in the Scleractinia/Corallimorpharia complex: implications for coral phylogeny. *Genome Biology and Evolution* **6(5)**:1086–1095 DOI 10.1093/gbe/evu084.

Liu H, Stephens TG, González-Pech RA, Beltran VH, Lapeyre B, Bongaerts P, Cooke I, Aranda M, Bourne DG, Forêt S, Miller DJ, van Oppen MJH, Voolstra CR, Ragan MA, Xin Chan C. 2018. Symbiodinium genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Communications Biology* **1**:95 DOI 10.1038/s42003-018-0098-3.

Low MEY, Sinniger F, Reimer JD. 2016. The order Zoantharia (Rafinesque, 1815) (Cnidaria, Anthozoa: Hexacorallia): supraspecific classification and nomenclature. *ZooKeys* **641**:1–80 DOI 10.3897/zookeys.641.10346.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27(6)**:764–770 DOI 10.1093/bioinformatics/btr011.

Martín-Durán JM, Vellutini BC, Marlétaz F, Cetrangolo V, Cvetesic N, Thiel D, Henriet S, Grau-Bové X, Carrillo-Baltodano AM, Gu W, Kerbl A, Marquez Y, Bekkouche N, Chourrout D, Gómez-Skarmeta JL, Irimia M, Lenhard B, Worsaae K, Hejnol A. 2021. Conservative route to genome compaction in a miniature annelid. *Nature Ecology and Evolution* **5**:231–242 DOI 10.1038/s41559-020-01327-6.

Metzger MJ, Paynter AN, Siddall ME, Goff SP. 2018. Horizontal transfer of retrotransposons between bivalves and other aquatic species of multiple phyla. *Proceedings of the National Academy of Sciences of the United States of America* **115(18)**:E4227–E4235 DOI 10.1073/pnas.1717227115.

Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I, Wong WY, Nowoshilow S, Kneitz S, Kawaguchi A, Fabrizius A, Xiong P, Dechaud C, Spaink HP, Volff J-N, Simakov O, Burmester T, Tanaka EM, Schartl M. 2021. Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* **5907845**:7845 DOI 10.1038/s41586-021-03198-8.

Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D. 2013. Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics* **14(2)**:193–202 DOI 10.1093/bib/bbs012.

Mizuyama M, Masucci GD, Reimer JD. 2018. Speciation among sympatric lineages in the genus *Palythoa* (Cnidaria: Anthozoa: Zoantharia) revealed by morphological comparison, phylogenetic analyses and investigation of spawning period. *PeerJ* **6**:e5132 DOI 10.7717/peerj.5132.

Montenegro J, Sinniger F, Reimer JD. 2015. Unexpected diversity and new species in the sponge-Parazoanthidae association in southern Japan. *Molecular Phylogenetics and Evolution* **89**:73–90 DOI 10.1016/j.ympev.2015.04.002.

Naville M, Henriet S, Warren I, Sumic S, Reeve M, Volff J-N, Chourrout D. 2019. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Current Biology* **29(7)**:1161–1168 DOI 10.1016/j.cub.2019.01.080.

Novák P, Neumann P, Macas J. 2020. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nature Protocols* **6(11)**:1325–1329 DOI 10.1038/s41596-020-0400-y.

Pantzartzi CN, Pergner J, Kozmik Z. 2018. The role of transposable elements in functional evolution of amphioxus genome: the case of opsin gene family. *Scientific Reports* **8(1)**:2506 DOI 10.1038/s41598-018-20683-9.

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**:526–528 DOI 10.1093/bioinformatics/bty633.

Paule J, von Döhren J, Sagorny C, Nilsson MA. 2021. Genome size dynamics in marine ribbon worms (Nemertea, Spiralia). *Genes* **12(9)**:1347 DOI 10.3390/genes12091347.

Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, Richards S, Niehuis O, Misof B. 2019. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Ecology and Evolution* **19(1)**:11 DOI 10.1186/s12862-018-1324-9.

Petri R, Brattås PL, Sharma Y, Jönsson ME, Pircs K, Bengzon J, Jakobsson J. 2019. LINE-2 transposable elements are a source of functional human microRNAs and target sites. *PLOS Genetics* **15(3)**:e1008036 DOI 10.1371/journal.pgen.1008036.

Poliseno A, Santos MEA, Kise H, MacDonald B, Quattrini AM, McFadden CS, Reimer JD. 2020. Evolutionary implications of analyses of complete mitochondrial genomes across order Zoantharia (Cnidaria: Hexacorallia). *Journal of Zoological Systematics and Evolutionary Research* **58(4)**:858–868 DOI 10.1111/jzs.12380.

Qiu Y, Köhler C. 2020. Mobility connects: transposable elements wire new transcriptional networks by transferring transcription factor binding motifs. *Biochemical Society Transactions* **48(3)**:1005–1017 DOI 10.1042/BST20190937.

Quattrini AM, Rodríguez E, Faircloth BC, Cowman PF, Brugler MR, Farfan GA, Hellberg ME, Kitahara MV, Morrison CL, Paz-García DA, Reimer JD, McFadden CS. 2020. Palaeoclimate ocean conditions shaped the evolution of corals and their skeletons through deep time. *Nature Ecology and Evolution* **4(11)**:1531–1538 DOI 10.1038/s41559-020-01291-1.

R Core Team. 2021. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/*.

Reimer JD, Takishita K, Ono S, Tsukahara J, Maruyama T. 2007. Molecular evidence suggesting interspecific hybridization in *Zoanthus* spp. (Anthozoa: Hexacorallia). *Zoological Science* **24**:346–359 DOI 10.2108/zsj.24.346.

Ronquist F, Huelsenbeck J. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574 DOI 10.1093/bioinformatics/btg180.

Santos MEA, Kise H, Fourreau CJL, Poliseno A, Pirro S, Reimer JD. 2023. The complete genome sequences of 13 species of Brachycnemina (Cnidaria, Hexacorallia, Anthozoa, Zoantharia). *Biodiversity Genomes* **2023**:10.56179/001c.72778 DOI 10.56179/001c.72778.

Santos MEA, Wirtz P, Montenegro J, Kise H, López C, Brown J, Reimer J. 2019. Diversity of Saint Helena Island and zoogeography of zoantharians in the Atlantic Ocean: Jigsaw falling into place. *Systematics and Biodiversity* **17(2)**:165–178 DOI 10.1080/14772000.2019.1572667.

Serrato-Capuchina A, Matute DR. 2018. The role of transposable elements in speciation. *Genes* **9(5)**:254 DOI 10.3390/genes9050254.

Shah A, Hoffman JI, Schielzeth H. 2020. Comparative analysis of genomic repeat content in gomphocerine grasshoppers reveals expansion of satellite DNA and helitrons in species with unusually large genomes. *Genome Biology and Evolution* **12(7)**:1180–1193 DOI 10.1093/gbe/evaa119.

Shoguchi E, Beedessee G, Hisata K, Tada I, Narisoko H, Satoh N, Kawachi M, Shinzato C. 2021. A new dinoflagellate genome illuminates a conserved gene cluster involved in sunscreen biosynthesis. *Genome Biology and Evolution* **13(2)**:evaa235 DOI 10.1093/gbe/evaa235.

Shoguchi E, Beedessee G, Tada I, Hisata K, Kawashima T, Takeuchi T, Arakaki N, Fujie M, Koyanagi R, Roy MC, Kawachi M, Hidaka M, Satoh N, Shinzato C. 2018. Two divergent *Symbiodinium* genomes reveal conservation of a gene

cluster for sunscreen biosynthesis and recently lost genes. *BMC Genomics* **19**:458 DOI 10.1186/s12864-018-4857-9.

Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, Takeuchi T, Hisata K, Tanaka M, Fujiwara M, Hamada M, Azadeh S, Fujie M, Usami T, Goto H, Yamasaki S, Arakaki N, Suzuki Y, Sugano S, Toyoda A, Kuroki Y, Fujiyama A, Medina M, Coffroth MA, Bhattacharya D, Satoh N. 2013. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology* **23**:1399–1408 DOI 10.1016/j.cub.2013.05.062.

Sinniger F, Chevaldonné P, Pawlowski JW. 2007. Mitochondrial genome of *Savalia savaglia* (Cnidaria, Hexacorallia) and early metazoan phylogeny. *Journal of Molecular Evolution* **64(2)**:196 DOI 10.1007/s00239-006-0015-0.

Sinniger F, Montoya-Burgos JI, Chevaldonné P, Pawloski J. 2005. Phylogeny of the order Zoantharia (Anthozoa, Hexacorallia) based on the mitochondrial ribosomal genes. *Marine Biology* **147**:1121–1128 DOI 10.1007/s00227-005-0016-3.

Smit AFA, Hubley R, Green P. 2013-2015. RepeatMasker Open 4.0.7. *Available at http://www.repeatmasker.org*.

Stampar SN, Broe MB, Macrander J, Reitzel AM, Brugler MR, Daly M. 2019. Linear mitochondrial genome in Anthozoa (Cnidaria): a case study in Ceriantharia. *Scientific Reports* **9**:6094 DOI 10.1038/s41598-019-42621-z.

Stelzer CP, Blommaert J, Waldvogel AM, Pichler M, Hecox-Lea B, Mark Welch DB. 2021b. 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 DOI 10.1186/s12915-021-01134-w.

Stelzer C-P, Pichler M, Hatheuer A. 2021a. Linking genome size variation to population phenotypic variation within the rotifer, *Brachionus asplanchnoidis*. *Communications Biology* **4(1)**:1–8 DOI 10.1038/s42003-020-01566-0.

Sun C, Shepard DB, Chong RA, López Arriaza J, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biology and Evolution* **4(2)**:168–183 DOI 10.1093/gbe/evr139.

Swift H. 1950. The constancy of desoxyribose nucleic acid in plant nuclei. *Proceedings of the National Academy of Sciences of the United States of America* **36(11)**:643–654 DOI 10.1073/pnas.36.11.643.

Talla V, Suh A, Kalsoom F, Dincă V, Vila R, Friberg M, Wiklund C, Backström N. 2017. Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (*Leptidea*) Butterflies. *Genome Biology and Evolution* **9(10)**:2491–2505 DOI 10.1093/gbe/evx163.

Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution* **9(4)**:678–687 DOI 10.1093/oxfordjournals.molbev.a040752.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* **17**:57–86.

Thomas CAJ. 1971. The genetic organization of chromosomes. *Annual Review of Genetics* **5**:237–256 DOI 10.1146/annurev.ge.05.120171.001321.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* **27(2)**:171–180 DOI 10.1111/j.1096-0031.2010.00329.x.

Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **2(5347605)**:102–105 DOI 10.1038/nature17951.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33(14)**:2202–2204 DOI 10.1093/bioinformatics/btx153.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**:973–982 DOI 10.1038/nrg2165.

Wickham H. 2016. *Ggplot2: elegant graphics for data analysis.* New York: Springer-Verlag.

Wingett SW, Andrews S. 2018. FastQ screen: a tool for multi-genome mapping and quality control. *F1000Research* **7**:71338 DOI 10.12688/f1000research.15931.2.

Wong WY, Simakov O, Bridge DM, Cartwright P, Bellantuono AJ, Kuhn A, Holstein TW, David CN, Steele RE, Martínez DE. 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 of the United States of America* **116(46)**:22915–22917 DOI 10.1073/pnas.1910106116.

Xia W, Li H, Cheng W, Li H, Mi Y, Gou X, Liu Y. 2020. High-quality genome assembly of *Chrysaora quinquecirrha* provides insights into the adaptive evolution of jellyfish. *Frontiers in Genetics* **11**:535 DOI 10.3389/fgene.2020.00535.

Yang S-Y, Bourgeois C, Ashworth CD, Reimer JD. 2013. *Palythoa* zoanthid 'barrens' in Okinawa: examination of possible environmental causes. *Zoological Studies* **52(1)**:39 DOI 10.1186/1810-522X-52-39.

Yu G. 2020. Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics* **69**:e96 DOI 10.1002/cpbi.96.

Yuan Z, Liu S, Zhou T, Tian C, Bao L, Dunham R, Liu Z. 2018. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genomics* **19(1)**:141 DOI 10.1186/s12864-018-4516-1.