BMC Genomics

RESEARCH

Open Access

Repeatome diversity in sea anemone genomics (Cnidaria: Actiniaria) based on the Actiniaria-REPlib library

Jeferson A. Durán-Fuentes[1,2*†], Maximiliano M. Maronna[1,3*†], Octavio M. Palacios-Gimenez[4,5,6], Elio R. Castillo[4,5,7], Joseph F. Ryan[8], Marymegan Daly[2] and Sérgio N. Stampar[1]

Background Genomic repetitive DNA sequences (Repeatomes, REPs) are widespread in eukaryotes, influencing biological form and function. In Cnidaria, an early-diverging animal lineage, these sequences remain largely uncharacterized. This study investigates sea anemone REPs (Cnidaria: Actiniaria) in a phylogenetic context. We sequenced and assembled *de novo* the genome of *Actinostella flosculifera* and analyzed a total of 38 nuclear genomes to create the first ActiniariaREP library (Actiniaria-REPlib). We compared Actiniaria-REPlib with Repbase and RepeatModeler2 libraries, and used dnaPipeTE to annotate REPs from genomic short-read datasets of 36 species for divergence landscapes.

Results Our study assembled and annotated the mitochondrial genomes, including 27 newly assembled ones. We re-annotated ~92% of the unknown sequences from the initial nuclear genome library, finding that 6.4–30.6% were DNA transposons, 2.1–11.6% retrotransposons, 1–28.4% tandem repeat sequences, and 1.2–7% unclassifiable sequences. Actiniaria-REPlib recovered 9.4x more REP sequences from actiniarian genomes than Dfam and 10.4x more than Repbase. It yielded 79,903 annotated TE consensus sequences (74,643 known, 5,260 unknown), compared to Dfam with 7,697 (3,742 known, 3,944 unknown) and Repbae (763 known).

Conclusions Our study significantly enhances the characterization of sea anemone repetitive DNA, assembling mitochondrial genomes, re-annotating nuclear sequences, and identifying diverse repeat elements. Actiniaria-REPlib vastly outperforms existing databases, recovering significantly more REP sequences and providing a comprehensive resource for future genomic and evolutionary studies in Actiniaria.

Keywords DnaPipeTE, Genome, Mobilome, Short-reads, Tandem repeat sequence, Transposable elements

†Jeferson A. Durán-Fuentes and Maximiliano M. Maronna contributed equally to this work.

*Correspondence:
Jeferson A. Durán-Fuentes
jeferson.duran-fuentes@unesp.br
Maximiliano M. Maronna
maxmaronna@gmail.com
[1] Laboratory of Evolution and Aquatic Diversity (LEDALab), São Paulo State University (UNESP), São Paulo, Bauru, Brazil
[2] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, OH, USA
[3] Institute of Oceanography, Federal University of Rio Grande, Rio Grande, Rio Grande Do Sul, Brazil
[4] Population Ecology Group, Institute of Ecology and Evolution, Friedrich Schiller University Jena, Jena E07743, Germany
[5] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstraße 4, Leipzig 04103, Germany
[6] Department of Organismal Biology – Systematic Biology, Evolutionary Biology Centre, Uppsala University, Uppsala SE-75236, Sweden
[7] Instituto de Biología Subtropical (IBS) CONICET-UNaM, Universidad Nacional de Misiones LQH, Posadas, Misiones, Argentina
[8] Whitney Laboratory for Marine Bioscience and the Department of Biology, University of Florida, Florida, USA

Introduction

Genomic content and repeatome diversity

Eukaryotic genomes present standard-universal traits related to form and function that have been inferred from cytogenetics and chromosome information, genomic kinetics (temperature-based genome DNA dissociation of base composition) [1, 2]) and genome size, based on the Feulgen Densitometry [3] and more recently based on Flow Cytometry [4]. Whole genome sequencing lets us access the nucleotide sequence level; combining nucleotide sequences with complementary information such as transcriptomics and gene expression, it is possible to describe and classify genomes with variable resolution (with some relevant caveats; e.g., see [5]). From broad-scale genome sequencing, it is possible to classify or compare genome structure criteria beyond classical euchromatin vs heterochromatin regions, such as coding vs non-coding regions, functional vs non-functional regions [6] and repetitive vs single-copy content, or even more specific ones, like repetitive expressed elements (mobilome), among others [7].

The combined insight of all of these perspectives provides a baseline for the expected, ancestrally shared structural aspects of the genome of animals [8, 9]. Most genomes present high numbers of repetitive DNA (repeatome, REP; [10]). Repetitive DNA may have different sequence structure and propagation strategies (Transposable elements (TEs) vs non-mobile sequence-only elements) and can be highly distributed as interspersed or tandem sequences (TEs vs satellite DNA) [11]. TEs constitute a substantial part of genomes in various organisms throughout the tree of life, accounting for over 45% of the human genome and up to 85% of the genome of maize [12]. The widespread presence of TEs is due to their ability to replicate through different mechanisms: retrotransposons (class I) copy and paste via an RNA intermediate, while most DNA transposons (class II) cut and paste within the host genome [13–15]. TEs are divided into autonomous elements, which encode proteins for transposition, and non-autonomous elements, which rely on the transposition machinery of autonomous counterparts for recognition [16]. Class I elements include short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), and long terminal repeat (LTR) retrotransposons. Class II elements consist of DNA transposons such as terminal inverted repeat (TIR) elements, Crypton, Helitron, and Maverick [17]. The transposition mechanism enables TEs to infiltrate the genome parasitically, often providing no benefit to the host organism [13]; however, examples highlight the beneficial roles that TEs can play in various organisms, contributing to adaptability, stress response, and overall survival in changing environments [18–21]. In other cases, TEs can cause harmful effects by triggering ectopic recombination, inducing chromosomal rearrangements, and disrupting coding sequences [22–24].

Another widely distributed repetitive element in eukaryotic genomes is satellite DNA (satDNA), consisting of tandemly arranged non-coding repetitive DNA primarily found in the centromeric and pericentromeric heterochromatin [25–27]. The evolution of satDNA is shaped by non-reciprocal genetic exchange mechanisms, including unequal crossing over, intra-strand homologous recombination, gene conversion, rolling-circle replication, and transposition; these processes can gradually increase the copy number of new sequence variants within a satDNA family across the genomes of a sexual population [25, 28–32]. Sequences within a satDNA family experience concerted evolution as repeat exchanges occur among family members through non-reciprocal genetic transfers between homologous and occasionally non-homologous chromosomes. The primary sequences of satDNAs tend to mutate rapidly, leading to distinct compositions and genomic distributions of satDNAs among strains, populations, subspecies, or species [25, 28, 30, 33–36]. However, there have been instances of satDNA sequence conservation over long evolutionary periods, as observed in several animal clades [37–41]. The library hypothesis suggests that species do not completely lose or gain specific satDNA lineages; instead, related species share a common repertoire of satDNAs that may independently increase or decrease in copy numbers during or after speciation [42]. Consequently, sequence divergence resulting from reproductive isolation can create species-specific profiles of satDNA sequence variants.

Due to their ability to propagate across genomes, sequences in the REP typically evolve much faster than single-copy DNA sequences. This, combined with their diversity and high dynamics, significantly complicates REP database construction and introduces biases into these databases. Repbase [43] and Dfam [44] are widely used reference databases for TE annotation, and combined with RepeatMasker [45], they identify repetitive sequences by searching the genome for homologous sequences present in the databases. The annotation of REPs remains a challenging yet essential task in genomics. Accurate annotation provides insights into the structural and functional complexities of genomes, potentially revealing how repetitive sequences contribute to evolutionary history and phenotypic diversity. Furthermore, understanding repetitive DNA is vital for comparative genomics, allowing researchers to identify conserved sequences and species-specific adaptations. As the number of genomes continues to rapidly grow, it has become increasingly clear that comprehensive repetitive DNA

annotations enhance our capacity to analyze and interpret genomic data effectively [46].

Cnidarian genomics with emphasis on Anthozoa

Because they are one of the early branching clades in the animal tree, cnidarians are a highly valuable group in studies of metazoan phylogenomics. Cnidaria represent ~12,500 valid species with three main groups: Anthozoa (anemones and corals, ~7,200 spp.), Medusozoa (jellyfishes including *Hydra*, ~4,120 spp.) and Endocnidozoa (myxozoans and kin, ~1,130 spp.) [47]. Taking into account the diversity of cnidarian genomes, Adachi et al. [48] analyzed genome sizes across Cnidaria, and Zhang and Jacobs [49] and Ying et al. [50] discussed methylation profiles related to genome evolution (see brief summary in Table 1). Within Cnidaria, studies typically focus on either clade Operculozoa (Medusozoa, Myxozoa, Polypodiozoa) or Anthozoa (Hexacorallia, Octocorallia). For Medusozoa, Santander et al. [51] reviewed current knowledge on genomics and recently Kon-Nanjo et al. [52] and Ahuja et al. [53] described hydrozoan genome sizes and REPs for *Hydra* and for species of order Siphonophorae, respectively. Comparative genomic analyses within the phylum Endocnidozoa have focused primarily on the genome evolution in relation to extreme reduction trends in species of Myxozoa and *Polypodium hydriforme*, including genomes sizes, protein-coding genes and number of orthologous gene groups [54–56].

Anthozoa has been the subject of a surge in genomic research, with over 150 genomes available in the NCBI-Assembly database [60–65]. Despite the availability of genomes for diverse octocorals, scleractinians, and actiniarian sea anemones, for most of these genomes, REP sections were not defined in detail and were not the main part of the results and discussion. One exception is the REP analysis led by Fourreau et al. [66], for Zoantharia. Perhaps unsurprisingly because REP are not fully annotated or deeply studied in cnidarian genomes, they are underrepresented in REP databases such as Dfam, which includes only eight species [44], and Repbase v29.03, which lists just one species [43]. Within the order Actiniaria, encompassing about ~1,200 valid species [47] and 53 genomic datasets available in the NCBI-Dataset ([60], accessed 11.15.2024), only *Anthopleura sola* Pearse & Francis, 2000 and *Nematostella vectensis* Stephenson, 1935 are represented in Dfam (Supplementary Table S1), and *N. vectensis* in Repbase.

From this context, we recognize (i) there are increasing numbers of genomes for anthozoans, including from high-quality sequencing techniques, (ii) there are highly diverse strategies for describing the content of these genomes, most of them with low emphasis on one of the most relevant parts of them (REPs), and (iii) low representation of these data and species in reference databases hampers more thorough study of cnidarian genome diversity and evolution. Consequently, here we endeavor to build a high quality REP database for selected Actiniaria species and use it to (i) to create Actiniaria-REPlib, a highly detailed REP library from 37 available Actiniaria genomic assemblies plus the de novo assembly of the genome of *Actinostella flosculifera* (Le Seuer, 1817); (ii) to compare alternative REP annotation pipelines and content of the 38 analyzed genomes of Actiniaria (Actiniaria-REPlib, RepBase, RepeatModeler2) based on their assemblies, (iii) compare the annotation and proportion of different classes of REPs in the 36 short reads datasets available for these species (several genomes assemblies did not have Illumina reads available; Table 2) available at NCBI using the Actiniaria-REPlib_v1 library, and (iii) to discuss strategies to enhance REP information quality in Anthozoa genomics. In the course of this work, we identify, assemble and annotate mitochondrial reads for those samples with no mitochondrial genome in the NCBI and use the mitogenomes to infer phylogenetic relationships that help interpret the structure and diversity of REPs in Actiniaria.

Table 1 Data summary for main cnidarian clades (Anthozoa, Medusozoa, and one main anthozoan clade (Actiniaria)). Data sources Santander et al. [51]; Animal Genome size Database [57], The Animal Chromosome Count database [58]. Genomes on a Tree database [59] and NCBI-datasets [60]. NA: Not available

Clade	Genome size (Megabases)			Chromosome Number		Repeatome (REP) (~ %)		Gene Number Count (~)	
	min	max	median	min	max	min	max	min	max
Anthozoa	286	1,142	649	18	54	30	50	18,425	62,650
Actiniaria	227	868	455	15	30	31	55	19,231	23,845
Medusozoa	263	3,567	711	12	40	27	64	17,200	66,150
Endocnidozoa	15	254	77	NA	NA	14	68	5,500	16,600

Table 2 Genome specifications for species used for construction (Const) and annotation (Annot) of the Actiniaria-REPlib_v1 library. Abbreviations– CVD: computationally very demanding; FC: Flow Cytometry; NA: not availablem; NGS: Next Generation Sequencing; tDNA: Mitogenomes used for phylogenetic analysis; SeqTech: Sequencing technologies (Illumina (I), PacBio (PB), and Oxford Nanopore (ONT)). '*': de novo assembly of mitogenomes deposited at NCBI. '**': de novo assembly of genome deposited at NCBI

Nr	Taxon	Genome size (technique) (Mb)	Scaffold N50 (count)	SeqTech	NCBI BioSample/ Reference	Assembly level	mtDNA	Const	Annot
Suborder Anenthemonae									
Superfamily Actinernoidea									
Family Actinernidae									
1	*Actinernus* sp.	NA (FC); 1,400 (NGS)	71.9 Mb (1,812)	PB + I	SAMN31231981	Scaffold	BK069892*	X	X
Superfamily Edwardsioidea									
Family Edwardsiidae									
2	*Edwardsia elegans*	NA (FC); 397 (NGS)	NA	I	SAMN43163413	Contig	PRJNA1247437*	X	X
3	*Nematostella vectensis*	0.34 (FC); 270 (NGS)	~ 17 Mb (47)	PB + I	SAMEA8534429	Chromosome	NC_008164.1	X	X
4	*Scolanthus callimorphus*	NA (FC); 596.8 (NGS)	~ 31 Mb (303)	PB + I	SAMN16376567	Chromosome	BK068674*	X	X
Suborder Enthemonae									
Superfamily Actinioidea									
Family Actiniidae									
5	*Anemonia viridis*	NA (FC); 401 (NGS)	2.1 kb (1,1 Mb)	I	SAMEA104356964	Scaffold	NC_037177	CVD	X
6	*Actinia equina*	NA (FC); 409 (NGS)	NA	PB	SAMN09602970	Contig	NA	X	NA
7	*Actinia mediterranea*	NA (FC); NA (NGS)	NA	I	SAMEA115283892	NA	BK069890*	NA	X
8	*Actinia tenebrosa*	NA (FC); 238.2 (NGS)	~ 188 kb (4,002)	I	SAMN10439458	Scaffold	NC_044902.1	X	X
9	*Actinostella flosculifera*	NA (FC); ~ 269 (NGS)	~ 3.1 kb (62,998)	I	SAMN45085772**	Scaffold	PV232310*	X	X
10	*Anthopleura artemisia*	NA (FC); ~ 342 (NGS)	15 Mb (1,169)	PB + Hi-C	SAMEA112465889	Chromosome	BK069895*	X	X
11	*Anthopleura elegantissima*	NA (FC); 322 (NGS)	~ 322 kb (4,216)	I + ONT	SAMN43844512	Scaffold	NA	X	NA
12	*Anthopleura sola*	NA (FC); 289 (NGS)	~ 10 Mb (269)	PB + I	SAMN24505220	Scaffold	BK068675*	X	X
13	*Anthopleura xanthogrammica*	NA (FC); 290 (NGS)	~ 14.8 Mb (133)	PB + Hi-C	SAMEA112465888	Chromosome	NA	X	NA
14	*Bunodosoma granuliferum*	NA (FC); 352 (NGS)	3.5 kb (278,782)	I	SAMN42720090	Scaffold	BK069896*	X	X
15	*Condylactis gigantea*	NA (FC); 239 (NGS)	~ 199.5 kb (4,656)	PB + I	SAMEA9267623	Chromosome	PRJNA1247437*	X	X
16	*Entacmaea quadricolor*	NA (FC); 428.3 (NGS)	~ 2.5 kb (249,586)	I	SAMN10992684	Scaffold	NC_049066.1	X	X
17	*Urticina crassicornis*	NA (FC); 302.1 (NGS)	~ 2.3 kb (188,453)	I	SAMN35990818	Scaffold	BK068676*	X	X
Family Actinodendridae									
18	*Actinodendron alcyonoideum*	NA (FC); 370 (NGS)	~ 7.5 kb (265,852)	I	SAMN42720097	Scaffold	BK069891*	X	X
19	*Actinodendron arboreum*	NA (FC); 628 (NGS)	1.5 kb (791,670)	I	SAMN42720085	Scaffold	BK069893*	CVD	X
Family Andvakiidae									
20	*Telmatactis stephensoni*	NA (FC); 485 (NGS)	NA	PB	SAMN27009947	Contig	NA	X	NA

Table 2 (continued)

Nr	Taxon	Genome size (technique) (Mb)	Scaffold N50 (count)	SeqTech	NCBI BioSample/ Reference	Assembly level	mtDNA	Const	Annot
Family Heteractidae									
21	*Heteractis aurora*	NA (FC); 248 (NGS)	7 kb (122.508)	I	SAMN42720084	Scaffold	BK069899*	X	X
22	*Heteranthus verruculatus*	NA (FC); 411 (NGS)	1.5 kb (481,366)	I	SAMN42720087	Scaffold	BK069900*	CVD	X
23	*Radianthus crispa*	NA (FC); ~275 (NGS)	~2.4 kb (166,207)	I	SAMN10992670	Scaffold	BK068678*	X	X
24	*Radianthus magnifica*	NA (FC); ~279 (NGS)	~2.8 kb (147,986)	I	SAMN10992683	Scaffold	BK068677*	X	X
Family Phymanthidae									
25	*Phymanthus crucifer*	NA (FC); ~297 (NGS)	~2.2 kb (315,387)	I	SAMN10246555	Scaffold	NC_027614.1	X	X
26	*Phymanthus loligo*	NA (FC); 320 (NGS)	2 kb (321,988)	I	SAMN42720083	Scaffold	BK069901*	X	X
Family Stichodactylidae									
27	*Stichodactyla helianthus*	NA (FC); ~297 (NGS)	~5.6 kb (209,545)	I	SAMN10992685	Scaffold	BK068679*	X	X
28	*Stichodactyla mertensii*	NA (FC); ~295 (NGS)	~5.5 kb (209,211)	I	SAMN10992686	Scaffold	BK068681*	X	X
29	*Stichodactyla tapetum*	NA (FC); 333 (NGS)	1.7 kb (375,204)	I	SAMN42720089	Scaffold	BK069903*	X	X
30	*Thalassianthus aster* (= *Stichodactyla* sp.)	NA (FC); 262 (NGS)	8.7 kb (128,901)	I	SAMN42720088	Scaffold	BK069902*	X	X
Superfamily Actinostoloidea									
Family Actinostolidae									
31	*Actinostola sp.*	NA (FC); 424 (NGS)	~383.1 kb (1,596)	PB	SAMN36377857	Scaffold	NA	X	NA
32	*Stomphia didemon*	NA (FC); ~158 (NGS)	~3.6 kb (76,020)	I	SAMN34510624	Scaffold	BK068680*	X	X
Superfamily Metridioidea									
Family Actinoscyphiidae									
33	*Actinoscyphia* sp.	NA (FC); 522 (NGS)	58.4 Mb (131)	PB + I	SAMN26810372	Scaffold	PRJNA1247437*	X	X
Family Aiptasiidae									
34	*Aiptasiogeton hyalinus*	NA (FC); 249 (NGS)	4 kb (211,325)	I	SAMN42720095	Scaffold	BK069894*	X	X
35	*Exaiptasia diaphana*	NA (FC); ~256 (NGS)	~442 kb (4,312)	I	SAMN03839803	Scaffold	NC_056771.1	X	X
Family Diadumenidae									
36	*Diadumene cincta*	NA (FC); 366 (NGS)	1.9 kb (425,998)	I	SAMN42720099	Scaffold	BK069897*	X	X
37	*Diadumene leucolena*	NA (FC); 360 (NGS)	1.6 kb (409,582)	I	SAMN42720098	Scaffold	BK069898*	X	X
38	*Diadumene lineata*	NA (FC); ~313 (NGS)	~17 Mb (137)	PB + I	SAMEA7536572	Scaffold	NC_045515.1	X	X
Family Hormathiidae									
39	*Paraphelliactis xishaensis*	NA (FC); 543 (NGS)	~761 kb (3,886)	PB + I	Feng et al., 2021	Scaffold	MT997141	X	X
Family Kadosactinidae									
40	*Alvinactis idsseensis*	NA (FC); 479 (NGS)	27.6 Mb (38)	Hi-C + I + ONT	Zhou et al., 2023	Chromosome	NA	X	NA

Table 2 (continued)

Nr	Taxon	Genome size (technique) (Mb)	Scaffold N50 (count)	SeqTech	NCBI BioSample/ Reference	Assembly level	mtDNA	Const	Annot
Family Metridiidae									
41	*Metridium farcimen*	NA (FC); ~ 339 (NGS)	~ 2.5 kb (209,616)	I	SAMN35990982	Scaffold	BK068682*	X	X
42	*Metridium senile*	NA (FC); ~ 390 (NGS)	~ 20 Mb (250)	PB + I	SAMEA110449715	Chromosome	HG423143.1	X	X

Results

Reads processing and assembly of genomes

We assembled the genome of *Actinostella flosculifera* from Illumina sequencing reads (Supplementary Table S2). We initially estimated genome size based on *k-mer* counting at 443 Mb; following trimming, we re-estimated total reads and bases to 265.47 million reads (90.2%) and 37.9 Gb (85.8%), respectively. We detected and removed 46.47 million of paired and unpaired reads (17.5%) and 6.7 Gb of bases (17.72%) containing exogenous DNA, resulting in "decontaminated" totals of 219 million reads (82.5%) and 31 Gb (82.3%), respectively (Supplementary Material S1 and Supplementary Table S2). We also removed the *A. flosculifera* mitogenome reads. The mitogenome is inferred to be circular and contain 19,504 bp (Supplementary Table S3). Following removal of the mitogenome reads, the genome size of *A. flosculifera* was estimated to be 261.1 Mb with a repeat content of approximately 84.26 Mb (32.3%), based on a *k-mer* (k = 21) analysis, (presumed diploid, heterozygosity of 1.5%: Supplementary Table S2). The best sub-optimal Platanus assembly was *k-mer* = 31, and this de novo genome assembly contained a N50 of 9,925 bp, BUSCO orthologs 66.77% and 25% (complete and partial), and genome size of ~ 268 Mb. Finally, after scaffolding with Ragtag (Supplementary Table S2), the assembly improved by 32% at N50 (13,099 bp) and 6.13% at BUSCO orthologs (70.55% complete and 21.1% partial), with a genome size of 269.4 Mb (Table 2, Supplementary Table S2).

The newly assembled and annotated mitogenomes comprise of *Actinernus* sp., *Actinia mediterranea* Schmidt, 1971, *Actinodendron alcyonoideum* (Quoy & Gaimard, 1833), *Actinodendron arboreum* (Quoy & Gaimard, 1833), *Actinoscyphia* sp., *Aiptasiogeton hyalinus* (Delle Chiaje, 1822), *Anthopleura artemisia* (Pickering in Dana, 1846), *A. sola*, *Bunodosoma granuliferum* (Le Sueur, 1817), *Condylactis gigantea* (Weinland, 1860), *Diadumene cincta* Stephenson, 1925, *Diadumene leucolena* (Verrill, 1866), *Edwardsia elegans* Verrill, 1869, *Heteranthus verruculatus* Klunzinger, 1877, *Metridium farcimen* (Brandt, 1835), *Phymanthus loligo* (Hemprich & Ehrenberg in Ehrenberg, 1834), *Radianthus crispa* (Hemprich & Ehrenberg in Ehrenberg, 1834), *Radianthus magnifica* (Quoy & Gaimard, 1833), *Scolanthus callimorphus* Gosse, 1853, *Stichodactyla* sp., *Stichodactyla helianthus* (Ellis, 1768), *Stichodactyla mertensii* Brandt, 1835, *Stichodactyla tapetum* (Hemprich & Ehrenberg in Ehrenberg, 1834), *Stomphia didemon* Siebert, 1973, and *Urticina crassicornis* (Müller, 1776).

Mitochondrial genomics and phylogenetic analysis

The length of the assembled mitogenomes varied from 15,969 to 20,910 bp (Supplementary Tables S4–5), with full conservation of gene order. The comparison of the aligned sequences and maximum likelihood (ML) phylogenomic reconstruction of the 36 actiniarian species used conserved positions of 13 protein-coding genes (PCGs) and 2 rRNAs concatenated of the mitogenomes (15,837 bp) (Supplementary Material S1) (*Actinia equina* (Linnaeus, 1758), *Actinostola* sp., *Alvinactis idsseensis* Zhou et al., 2023, *Anthopleura elegantissima* (Brandt, 1835), *Anthopleura xanthogrammica* (Brandt, 1835), and *Telmatactis stephensoni* Carlgren, 1950 were not included in these analyses because they do not have short reads available at NCBI; see Table 2). Maximum-likelihood phylogenetic analyses showed high support in most branches (Figs. 1 and 2). We recovered suborder Enthemonae as a monophyletic group with high support (SH-aLRT = 100%/parametric aLRT = 1/aBayes test = 1/ultrafast bootstrap = 100%). Within this suborder, we found that superfamily Actinioidea is more closely related to Metridioidea than to Actinostoloidea (*S. didemon*) with 81%/1/1/80% support. The suborder Anenthemoneae is represented by members of superfamilies Edwardsioidea (*E. elegans*, *N. vectensis*, and *S. callimorphus*) and Actinernoidea (*Actinernus* sp.) (100%/1/1/100%); this subfamily is monophyletic and sister to (Actinostoloidea (Actinioidea, Metridiodea)) (Figs. 1 and 2).

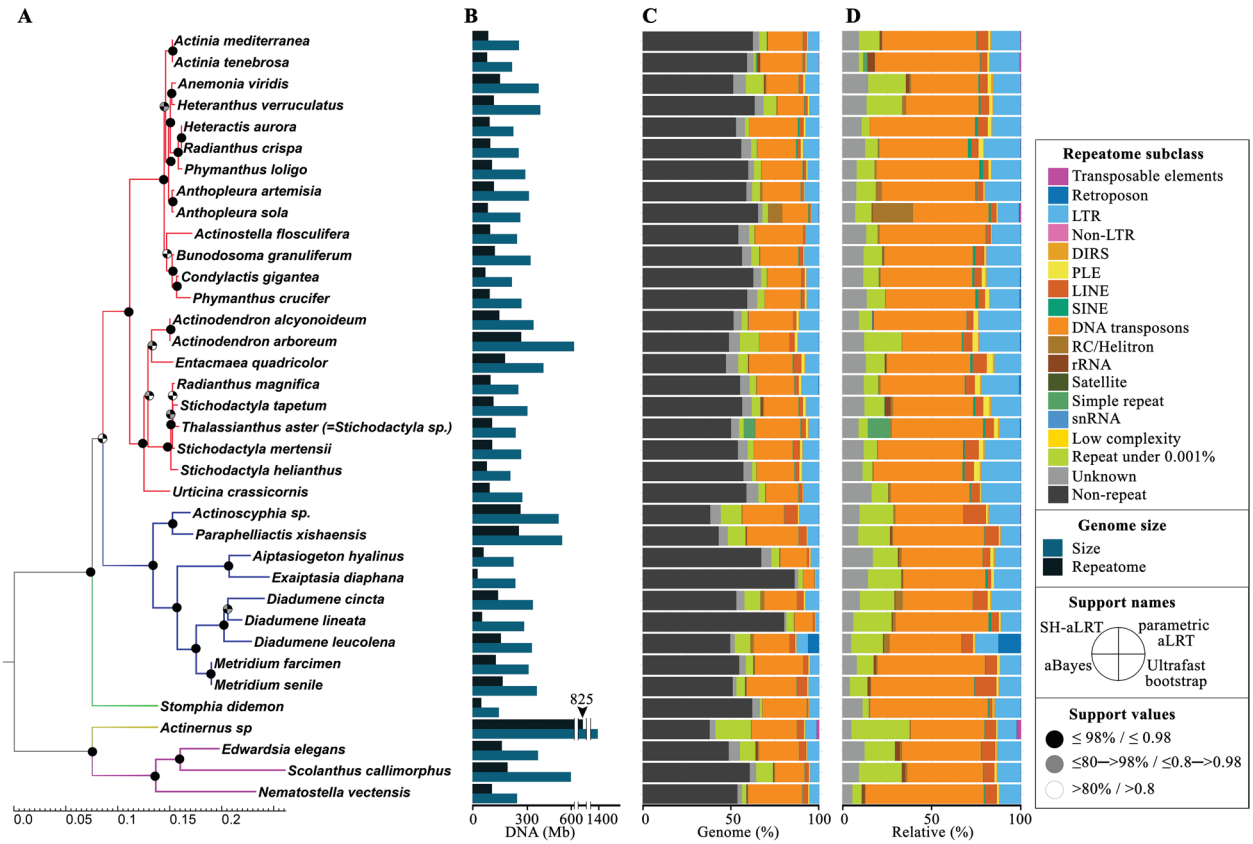

Fig. 1 Annotation and comparison of 36 actiniarian genomes using the Actiniaria-REPlib_v1 libraries in dnaPipeTE pipeline. **A** Phylogenetic reconstruction based on maximum likelihood analysis using the concatenated mitogenome dataset (13 protein-coding genes and rRNA genes); **B** genome and REP size; **C** repeat class abundance; and **D** relative percentage of repeat class abundance of the REP. Superfamilies: Actinernoidea (light brown branch), Actinioidea (red branch), Actinostoloidea (green branch), Edwardsioidea (purple branch), and Metridioidea (blue branch)

Construction of the Actiniaria-REPlib library

Initially, 42 Actiniaria genomes were included for construction of the Actiniaria-REPlib library, but four of them (*A. mediterranea* Schmidt, 1971, *Anemonia viridis* (Forsskål, 1775), *A. arboreum* (Quoy & Gaimard, 1833), and *H. verruculatus* Klunzinger, 1877 were excluded from the analyses because they do not have assembled genomes available at NCBI; see Table 2) proved to be computationally demanding when using RepeatModeler2, due to the Scaffold N50 and count being 1.5–2.1 kb and ~0.48–1.1 Mb, respectively. Furthermore, we performed comparative analyses between the newly assembled genome of *A. flosculifera* and the 37 other actiniarian genome assemblies obtainable from the NCBI database (Supplementary Table S6). These species represent five superfamilies, 15 families, and 26 genera, and have genomes that range in size from 0.16 to ~1.4 Gb. Three of these species have fragmented assemblies organized in contigs, 29 in scaffolds, and six have chromosome-level assemblies (*A. xanthogrammica*, *A. idsseensis*, *C. gigantea*, *N. vectensis*, *Metridium senile*, and *S. callimorphus*) (Supplementary Table S6).

The initial construction of the Actiniaria library (Actiniaria-REPlib_A) included main types of TEs (DNA, LINE, LTR, PLE, RC, and SINE) and tandem repeat (TR) sequences (rRNA, snRNA, satellite DNA, simple repeat, among others) (Fig. 3). Among the 38 REP libraries of Actiniaria, we found the greatest number of REP sequences in *Entacmaea quadricolor* (Leuckart in Rüppell & Leuckart, 1828), which contains 5,429 REP sequences, comprising 188 for DNA transposons (~ 3.5%), 632 for retrotransposons (~ 11.6%), 16 for TRS (~ 0.3%), and 4,593 unknown REP sequences (~ 84.3%). The merger of the 38 REP libraries contains 126,474 REP sequences, 4,637 of which are DNA transposons (~ 3.7%), 11,346 are retrotransposons (~ 9%), 541 are TRS (~ 0.43%), and 109,950 are unknown REP sequences (~ 86.9%) (Supplementary Table S6). Actiniaria-REPlib_B contains 79,903 REP sequences, which reflects a reduction of 36.85% in the number of redundant sequences compared to the initial combined database. It contains

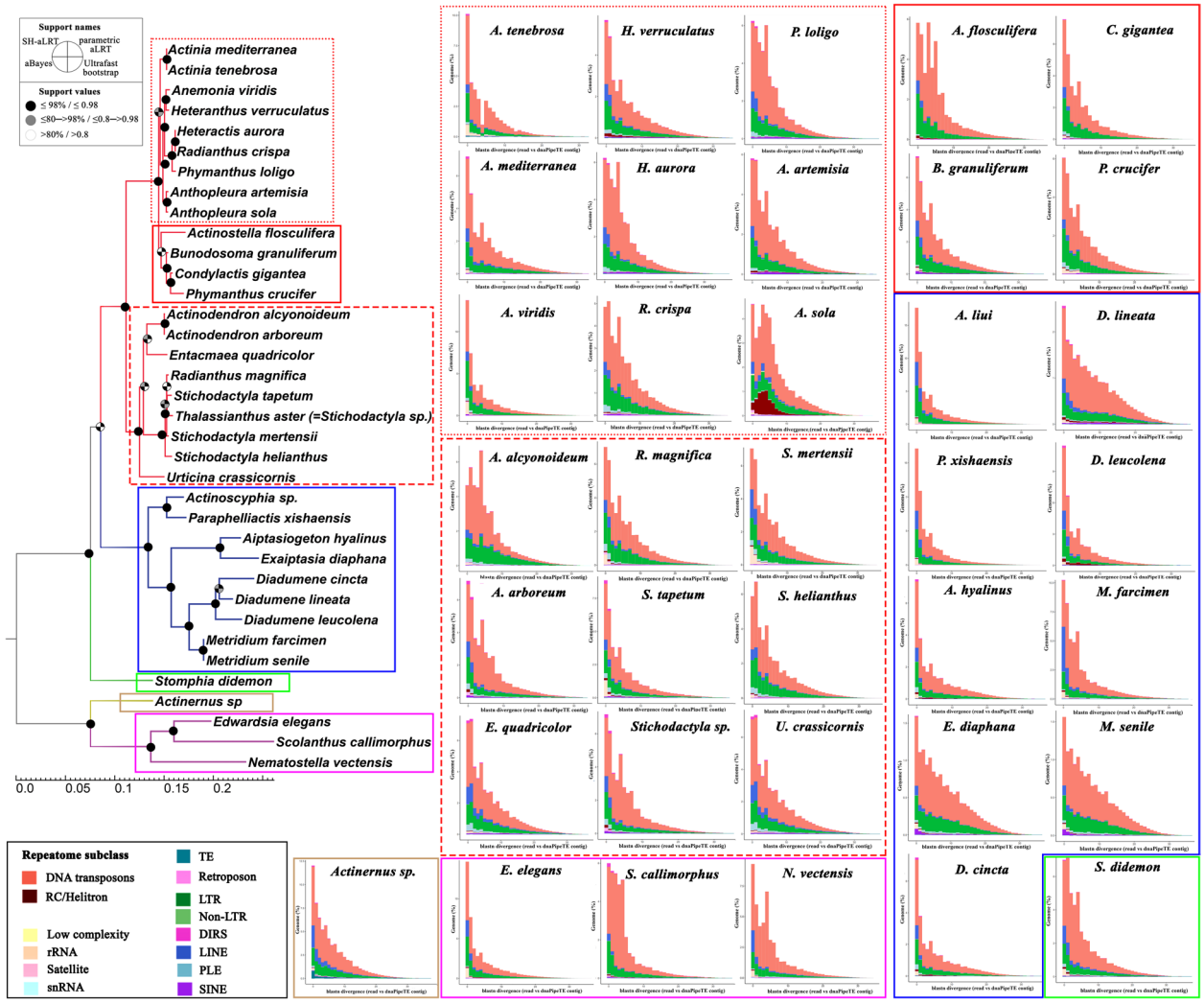

Fig. 2 Transposable element divergence landscapes for 36 species of actiniarians. Superfamilies: Actinernoidea (light brown branch), Actinioidea (red branch), Actinostoloidea (green branch), Edwardsioidea (purple branch), and Metridioidea (blue branch)

13,604 annotated sequences: 3,833 DNA transposons (~ 4.8%), 9,520 retrotransposons (~ 11.9%), 251 TRS (~ 0.3%), and 66,299 unannotated REP sequences (~ 83%) (Supplementary Table S7). We used the nomenclature level 1/level 2-level 3 when naming REPs (see below for more details).

The unknown REP sequences in Actiniaria-REPlib_B were re-annotated through DeepTE, TEsorter, TEclass2, and DANTE. DeepTE identified 92% (61,006) of the unknown REP sequences, classifying 41,258 (62.2%) as DNA transposons, 19,748 (29.8%) as retrotransposons, and failing to classify 5,293 (8%) (Supplementary Table S8). TEsorter re-annotated less than 1% (379; 0.57%) of the same unknown REP dataset: 38 DNA transposons and 341 retrotransposons (Supplementary Table S9). We examined the overlap in annotations between DeepTE and TEsorter and found 346 that were annotated by both programs. Of these 346, 265 had conflict in classification (e.g., rnd- 1_Actinernus_sp- 1232 was re-annotated in DeepTE as DNA/TcMar and in TEsorter as LINE) (Supplementary Table S10). TEclass2 classified 246 of these 265 conflicting sequences as 59 DNA transposons and 187 retrotransposons (Supplementary Table S11). DANTE only classified 236 of the conflicting sequences, 28 DNA transposons and 208 retrotransposons (Supplementary Table S12). We next used TEclass2 and DANTE to resolve 196 of these 265 sequences with conflicting annotation (3 DNA transposons and 193 retrotransposons), and the remaining 69 sequences were annotated as TE-level (Transposable element) (Supplementary Table S11–13). Actiniaria-REPlib_v1 library contains 79,903 REP sequences, 45,052 of which are DNA transposons (~ 56.4%), 29,340 are retrotransposons (~ 36.8%), 251 are TRS (~ 0.3%),

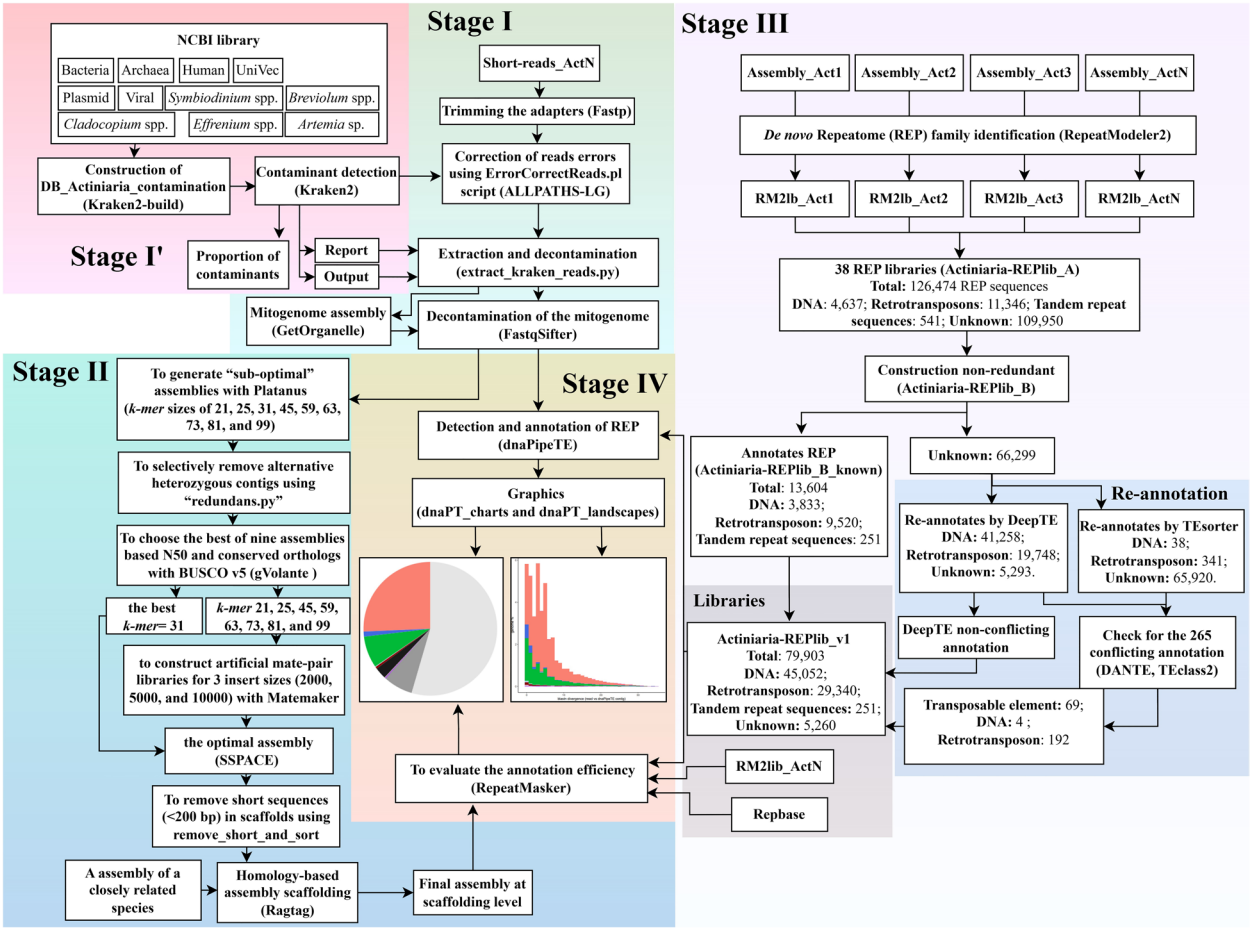

Fig. 3 Actiniaria-REPlib pipeline– Stage I: sequencing data pre-processing; Stage I': exogenous DNA removal; Stage II: protocol for genome assembly using Illumina sequences; Stage III: de novo construction of the Actiniaria-REPlib_v1; Stage IV: quantification of the repeatome (REP) content. Abbreviation– RM2 lib: RepeatModeler2 output/library; LTR: long terminal repeat; LINE: long interspersed nuclear element; PCG: protein-coding genes; PLE: Penelope-like element; SINE: short interspersed nuclear element

and 5,260 are unknown REP sequences (~ 6.6%) (Supplementary Table S14). Likewise, we have managed to re-annotate ~92% of the unknown sequences of the Actiniaria-REPlib_B library (from 66,299 to 5,260 unknown sequences).

Classification of the "Actiniaria-REPlib" library

We classified sequences within Actiniaria-REPlib library into four levels following Liu et al. [67], modifying this to differentiate LTR and Non-LTR Retrotransposons, and tandem repeat sequences (TRs) at the level of Type, and RNA and simple sequence repeats (SSRs) at the level of Class (Supplementary Table S14). RepeatMasker.lib (Repbase's default reference data) uses the nomenclature of Liu et al. [67] to generate the de novo annotation by RepeatModeler at the 3 different levels, these are coded as (i) level 1/level 2-level 3 (e.g., DNA/Crypton-A), (ii) level 1/level 2 (e.g., LTR/Copia), or (iii) level 1 (e.g., PLE) (Supplementary Table S6). Level 2 is encoded as the superfamily level and level 3 as the clade level. We formatted our sequence annotations from DeepTE, DANTE and TEclass2 to adopt this convention (e.g., from ClassI_LTR_BEL to LTR/Bel-Pao; Supplementary Table S7–13). We included "Retroposon" as a category rather than following Lui et al. [67] to distinguish betw een "LINE", "LTR", "DIRS", "PLE", and "SINE" because DeepTE and TEclass2 were not able to annotate any of the five classes of retrotransposon classes. Similarly, for those cases where no classification was defined by either tool, we included "TE" (Transposable element) as final annotation definition. Doing so, Actiniaria-REPlib contains 49 superfamilies of TE and three of TRs, and 58 clades of TE.

Quantification and annotation of the REPs using "Actiniaria-REPlib" library

We characterized the repetitive DNA content in the actinarian genome assemblies using homology-based and de novo approaches. To measure the effect of annotating REPs of actinarian genomes using Actiniaria-REPlib rather than more general REP libraries like Repbase and RM2 lib using RepeatMasker, we compared the number of identified repetitive elements across libraries (Fig. 4). As expected, Actiniaria-REPlib library identified many more repetitive elements in all assemblies as compared to the RM2 lib and Repbase libraries. The average percentage of REP sequences identified using Actiniaria-REPlib was 48.2% with a standard deviation of 9.4%, while RM2 lib and Repbase identified an average of 8.4% (± 3.6) and 7.8% (± 3.5%), respectively (Fig. 4, Table 3, and Supplementary Table S15).

When using the Actiniaria-REPlib, DNA transposons are inferred to be the most common repeat masked in the genomes (28.8 ± 6.3%), followed by long-terminal repeats (LTRs, 10.1 ± 2.1%) (Table 3 and Supplementary Table S15). Analyses of repeat content in the actiniarian genomes (except *A. equina*, *Actinostola* sp., *A. idsseensis*, *A. elegantissima*, *A. xanthogrammica*, and *T. stephensoni*) based on low-coverage sequencing reads (0.25 × genome coverage). The Actiniaria-REPlib library as custom database for annotation in dnaPipeTE [68] the total of REP contents for these species were estimated to be 13.9–62% (43.7 ± 9.3%) (Fig. 1C and Table 3). As with the assemblies, DNA transposons were the most common repeats at 6.4–30.6% (21.3 ± 5.1%), followed by LTRs with 2.1–11.6% (7 ± 2.1%), TR sequences with 1.4–28.40.5%, and unclassifiable sequences with 1.2–7% (Fig. 1D and Table 3).

Discussion and conclusion

Mitochondrial genomes and Actiniaria phylogeny

This study includes 27 Actiniaria mitogenomes (Fig. 1 and 2, Supplementary Tables S3 and S5) that were not available in the NCBI database. The primary use of the mitogenomes in this study was as a source of phylogenetic information. The results are very similar to those reported

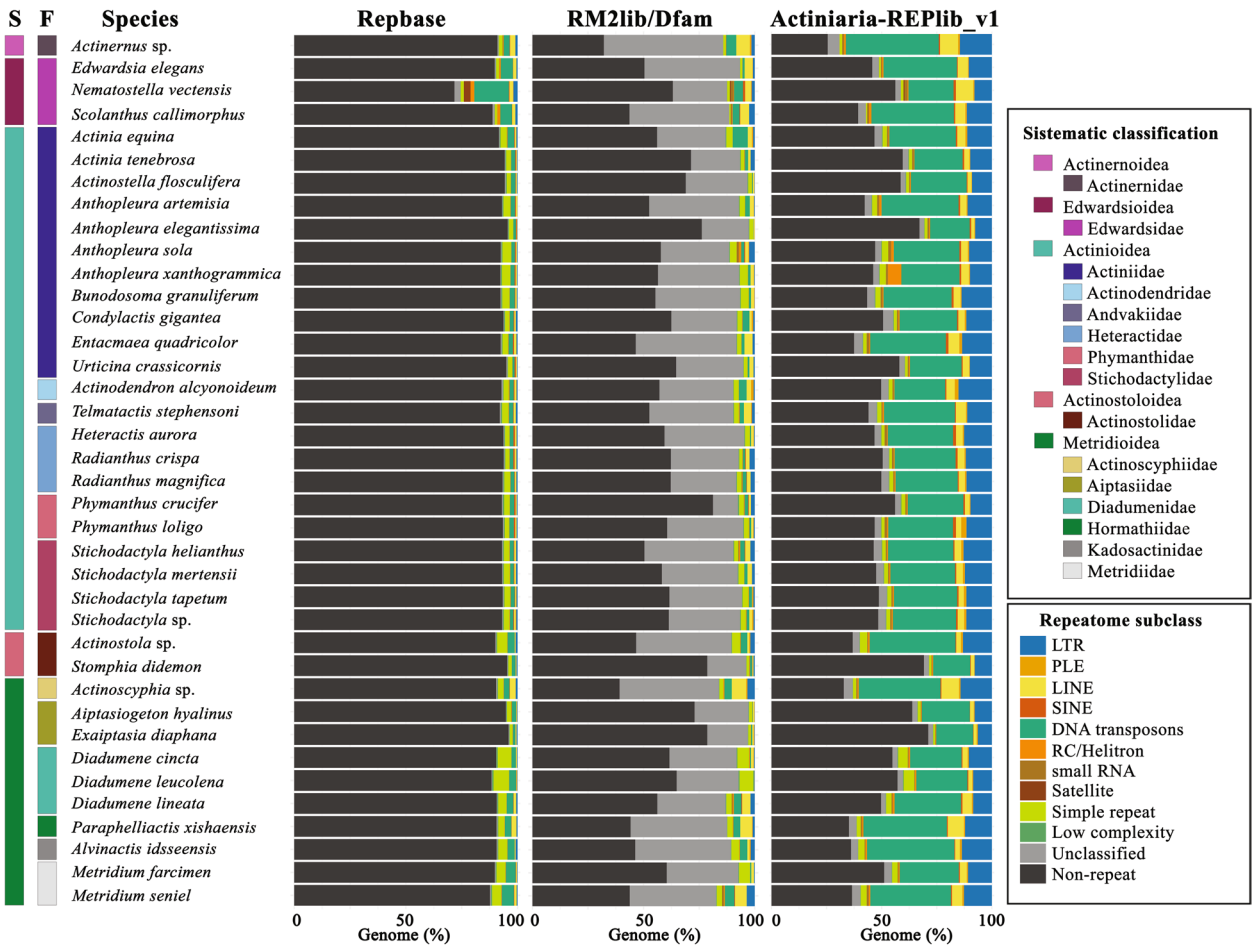

Fig. 4 Comparison of the 38 annotation genomes based on three libraries of REPs using RepeatMasker

Table 3 Efficiency in the annotation of three libraries of REPs (Repbase, the library built by RepeatModeler2 of each genome (RM2lib), and Actiniaria-REPlib for 38 actinarian genomes). Colors in the column for species represent their superfamilial taxonomic classification – light red: Edwardsioidea; light purple: Actinioidea; light green: Actinostoloidea; light orange: Metridioidea. Abbreviations– DNAt: DNA transposons; RT: Retrotransposons; REP: Total repeatome; TRs: Tandem repeat sequences

Nr	Species	Database	Genome (%)					
			RT	DNAt	TRs	REP	Unknown	Non-repeat
1	*Actinernus sp.*	Repbase	3.6	3.0	2.0	8.5	0.50	90.97
		RM2lib	8.3	4.7	1.2	14.2	53.67	32.12
		Actiniaria-REPlib	24.1	43.0	2.0	69.2	5.33	25.50
2	*Edwardsia elegans*	Repbase	2.1	6.1	1.6	9.8	0.45	89.76
		RM2lib	4.6	1.1	1.1	6.7	42.94	50.39
		Actiniaria-REPlib	15.9	34.0	1.5	51.3	2.98	45.73
3	*Nematostella vectensis*	Repbase	3.9	17.2	4.5	25.6	2.69	71.72
		RM2lib	5.3	4.2	3.0	12.5	24.32	63.15
		Actiniaria-REPlib	19.0	22.8	3.4	45.1	2.64	61.20
4	*Scolanthus callimorphus*	Repbase	2.5	6.5	1.3	10.3	0.90	88.79
		RM2lib	6.8	3.5	1.2	11.5	44.83	43.64
		Actiniaria-REPlib	17.2	38.6	1.4	57.2	3.47	39.33
5	*Actinia equina*	Repbase	1.5	3.3	3.2	8.0	0.30	91.66
		RM2lib	3.2	6.8	3.1	13.1	30.83	56.06
		Actiniaria-REPlib	16.4	30.5	3.0	49.9	3.39	46.74
6	*Actinia tenebrosa*	Repbase	1.0	2.0	2.5	5.5	0.28	94.23
		RM2lib	3.0	1.4	2.2	6.6	22.04	71.32
		Actiniaria-REPlib	13.3	22.1	2.5	37.9	2.53	59.54

Table 3 (continued)

7	*Actinostella flosculifera*	Repbase	1.0	2.1	2.2	5.3	0.47	94.25
		RM2lib	0.8	0.5	2.0	3.3	27.72	68.97
		Actiniaria-REPlib	11.5	25.7	2.0	39.2	2.25	58.59
8	*Anthopleura artemisia*	Repbase	1.1	2.1	3.5	6.7	0.24	93.02
		RM2lib	2.2	2.1	2.8	7.1	40.38	52.49
		Actiniaria-REPlib	15.1	36.0	3.5	54.6	3.15	42.28
9	*Anthopleura elegantissima*	Repbase	0.8	1.2	2.2	4.2	0.23	95.58
		RM2lib	0.5	0.2	2.0	2.7	21.13	76.18
		Actiniaria-REPlib	10.1	18.4	2.3	30.8	2.10	67.15
10	*Anthopleura sola*	Repbase	1.2	1.7	4.3	7.2	0.33	92.45
		RM2lib	4.7	2.4	4.5	11.5	30.73	57.74
		Actiniaria-REPlib	14.8	30.9	4.5	50.2	2.77	47.03
11	*Anthopleura xanthogrammica*	Repbase	1.3	1.9	4.2	7.4	0.28	92.28
		RM2lib	1.9	1.2	3.9	7.1	36.47	56.47
		Actiniaria-REPlib	14.7	32.4	4.1	51.2	2.66	46.14
12	*Bunodosoma granuliferum*	Repbase	1.3	2.3	3.7	7.3	0.38	92.32
		RM2lib	1.7	1.1	3.9	6.6	38.05	55.31
		Actiniaria-REPlib	18.3	31.3	3.5	53.1	3.46	43.40
13	*Condylactis gigantea*	Repbase	1.8	1.9	2.3	6.0	0.30	93.68
		RM2lib	2.4	3.4	2.4	8.1	29.39	62.47
		Actiniaria-REPlib	15.8	26.6	2.4	44.8	4.58	50.63

Table 3 (continued)

14	*Entacmaea quadricolor*	Repbase	2.1	2.1	2.9	7.1	0.47	92.41
		RM2lib	4.6	1.5	2.2	8.2	45.32	46.46
		Actiniaria-REPlib	20.9	34.7	3.0	58.6	3.93	37.47
15	*Urticina crassicornis*	Repbase	1.3	1.0	2.2	4.6	0.47	94.96
		RM2lib	2.4	0.9	1.8	5.1	30.22	64.64
		Actiniaria-REPlib	13.8	23.9	1.9	39.6	2.41	58.02
16	*Actinodendron alcyonoideum*	Repbase	1.8	2.2	2.6	6.6	0.65	92.79
		RM2lib	3.7	3.6	2.4	9.7	33.23	57.11
		Actiniaria-REPlib	21.2	23.3	2.5	47.0	3.27	49.76
17	*Telmatactis stephensoni*	Repbase	2.0	2.1	3.4	7.4	0.55	92.02
		RM2lib	4.9	2.3	2.5	9.6	37.79	52.58
		Actiniaria-REPlib	16.8	32.8	2.7	52.3	3.72	44.02
18	*Heteractis aurora*	Repbase	1.9	1.8	2.3	6.0	0.38	93.66
		RM2lib	1.6	0.5	2.5	4.7	35.95	59.40
		Actiniaria-REPlib	17.6	29.9	2.8	50.3	2.91	46.76
19	*Radianthus crispa*	Repbase	1.9	1.7	2.2	5.9	0.36	93.79
		RM2lib	4.3	1.1	1.8	7.3	30.46	62.26
		Actiniaria-REPlib	16.5	28.1	2.2	46.8	2.75	50.45
20	*Radianthus magnifica*	Repbase	1.5	1.8	3.1	6.4	0.37	93.26
		RM2lib	3.7	2.0	2.6	8.3	29.47	62.19
		Actiniaria-REPlib	15.5	28.7	2.6	46.8	3.38	49.84

Table 3 (continued)

21	*Phymanthus crucifer*	Repbase	1.5	2.0	3.0	6.6	0.47	92.97
		RM2lib	3.0	1.8	2.8	7.6	11.21	81.21
		Actiniaria-REPlib	12.9	25.7	2.7	41.4	2.58	56.07
22	*Phymanthus loligo*	Repbase	1.7	2.1	2.3	6.1	0.47	93.42
		RM2lib	1.6	1.1	2.5	5.2	34.25	60.60
		Actiniaria-REPlib	17.7	29.9	2.8	50.3	2.86	46.82
23	*Stichodactyla helianthus*	Repbase	1.7	1.9	2.9	6.5	0.42	93.04
		RM2lib	4.5	2.5	2.6	9.6	40.03	50.38
		Actiniaria-REPlib	17.5	30.1	2.4	50.0	3.61	46.36
24	*Stichodactyla mertensii*	Repbase	1.7	1.7	3.1	6.6	0.36	93.09
		RM2lib	3.3	1.5	2.9	7.7	34.14	58.21
		Actiniaria-REPlib	16.7	29.8	2.6	49.1	3.38	47.48
25	*Stichodactyla tapetum*	Repbase	1.3	2.0	2.9	6.2	0.43	93.34
		RM2lib	1.3	1.5	3.0	5.8	32.61	61.62
		Actiniaria-REPlib	15.8	29.2	2.7	47.7	3.59	48.73
26	*Thalassianthus aster* (=*Stichodactyla sp.*)	Repbase	1.9	1.7	3.0	6.6	0.34	93.10
		RM2lib	2.5	1.5	2.7	6.6	32.05	61.33
		Actiniaria-REPlib	16.3	29.2	2.7	48.1	3.44	48.42

Table 3 (continued)

27	*Actinostola sp.*	Repbase	1.3	3.3	4.9	9.5	0.42	90.04
		RM2lib	3.4	3.2	3.9	10.5	42.86	46.64
		Actiniaria-REPlib	16.5	39.7	3.9	60.1	3.09	36.84
28	*Stomphia didemon*	Repbase	0.9	1.7	1.7	4.3	0.31	95.35
		RM2lib	1.2	1.2	1.5	3.9	17.48	78.64
		Actiniaria-REPlib	9.9	17.2	1.6	28.6	2.19	69.17
29	*Actinoscyphia* sp.	Repbase	3.6	2.7	2.8	9.1	0.50	90.40
		RM2lib	10.3	3.8	2.0	16.1	44.69	39.19
		Actiniaria-REPlib	23.3	37.7	2.2	63.2	4.04	32.79
30	*Aiptasiogeton hyalinus*	Repbase	0.6	2.0	2.4	5.1	0.17	94.78
		RM2lib	0.8	0.3	1.7	2.8	24.25	72.93
		Actiniaria-REPlib	10.0	22.1	1.7	33.8	2.32	63.86
31	*Exaiptasia diaphana*	Repbase	1.0	1.3	1.5	3.7	0.29	95.97
		RM2lib	1.1	0.7	1.3	3.1	18.25	78.64
		Actiniaria-REPlib	8.5	17.0	1.2	26.8	2.12	71.12
32	*Diadumene cincta*	Repbase	0.8	2.1	6.6	9.5	0.11	90.43
		RM2lib	1.7	0.9	5.7	8.2	30.16	61.64
		Actiniaria-REPlib	13.8	24.0	5.0	42.8	2.30	54.87

Table 3 (continued)

33	*Diadumene leucolena*	Repbase	0.7	3.2	7.6	11.4	0.26	88.32
		RM2lib	0.3	0.4	6.7	7.3	27.81	64.87
		Actiniaria-REPlib	11.1	23.8	5.5	40.4	2.44	57.18
34	*Diadumene lineata*	Repbase	1.9	3.1	4.1	9.1	0.21	90.66
		RM2lib	6.0	3.9	3.2	13.1	30.73	56.17
		Actiniaria-REPlib	14.0	31.0	3.2	48.2	2.05	49.72
35	*Paraphelliactis xishaensis*	Repbase	2.8	3.0	3.2	9.0	0.47	90.56
		RM2lib	6.5	3.4	2.6	12.5	43.32	44.14
		Actiniaria-REPlib	20.5	38.4	2.6	61.5	3.39	35.14
36	*Alvinactis idsseensis*	Repbase	1.3	3.3	4.4	9.0	0.40	90.61
		RM2lib	3.3	3.6	3.8	10.8	43.01	46.24
		Actiniaria-REPlib	17.0	40.3	3.4	60.8	3.13	36.09
37	*Metridium farcimen*	Repbase	0.9	4.5	4.5	9.8	0.37	89.80
		RM2lib	1.5	0.6	5.3	7.4	32.20	60.41
		Actiniaria-REPlib	15.0	27.4	3.1	45.5	3.38	51.13
38	*Metridium senile*	Repbase	1.7	5.6	4.7	11.9	0.47	87.66
		RM2lib	9.2	4.7	3.5	17.4	38.89	43.75
		Actiniaria-REPlib	18.6	37.3	3.9	59.7	3.77	36.53

by other studies, including those based on conventional datasets of nuclear and mitochondrial markers [69–75] as well as those based on genome-scale data like UCEs [76–78]. This broad congruence contravenes expectations of discordance between signal from mitochondrial and nuclear genes (reviewed in Quattrini et al. [78]). The one notable difference is that in our tree, the actinostoloidean *Stomphia* is a sister group of the superfamily Actinioidea and Metridioidea, which is in contrast to recent studies based on genome-scale data [76, 77]. Because our study includes only one member of Actinostoloidea, it cannot address the monophyly of that group, but we think it noteworthy that our topology recalls those from studies that have found a polyphyletic Actinostoloidea [69, 74].

On repeatome (REP) access, description and basic annotation

Genomic annotation uses similarity between sequences of known function and identity to predict function and identity of unknown sequences, and so depends in large part on the quality and depth of previous knowledge that can be used to build predictions related to a particular content (in this case, databases as guiding references are fundamental). It is common for REP characterization to be absent from, or incomplete in many genome publications. This can be attributed to the limited scope of individual studies, computational time required for analysis, and/or the limited utility of existing reference databases for a particular genome, among other factors. There are pros and cons to each strategy for annotating the REP, especially analyzing short read data. The repetitive nature of genomes makes the assembly step difficult, and subsequent correctness of REP annotation will vary (usually, it will present an incomplete set of repetitive elements; [5]); on the other hand, short reads present a higher amount of information but are difficult to process and to relate to general genome content. The REP analysis by Fourreau et al. [66] in Zoantharia highlights this problem: the analyses offer important new insights and identify a large number of repeats, but contain a large number of unknowns and no final classification. This may reflect the underlying short read data, the limits of the comparative database, or both of these issues, with varying impacts across species and genomes. Another relevant issue is REP deposit details in major public repositories, like the International Nucleotide Sequence Database Collaboration (INSDC; [79]). In one of its main sections, the NCBI [60] defines traditional gene content but does not define REP in similar detail: "Coding regions (CDS) and RNAs, such as tRNAs and rRNAs, must have a corresponding gene feature. However, other features such as repeat_regions and misc_features do not have a corresponding gene or locus_tag." [80].

Given this context, we recommend that priorities be developed for genomics research. REP characterization would benefit from several community-driven actions: (i) improvement in deposit formatting, as stated previously by Santander et al. [51] and Brown et al. [81]; (ii) improvement in and explicit documentation of curation (see Goubert et al. [82] and Peona et al. [83]); (iii) experimental validation; and (iv) enhancement of strategies to standardize comparative approaches to REP classification, such as inclusion of TE-classification within the Genomes Standards Consortium Minimum Information about a Genome Sequence (MIGS) and Minimum Information about any (x) Sequence (MIxS) [84, 85].

Actiniaria-REPlib, Actiniaria and REPs´s DBs

Actiniaria-REPlib recovered 9.4 ×more REP sequences from actiniarian genomes than Dfam and 10.4 ×more than Repbase. It yielded 79,903 annotated TE consensus sequences (74,643 known, 5,260 unknown; 38 sea anemones species), compared to Dfam v3.8 (3,742 known, 3,944 unknown; 8 cnidarian species) and Repbase (763 known; *N. vectensis*) (Supplementary Table S1). Additionally, it led to a 5.2x (median) ± 1.7x (SD) increase in annotations compared to Repbase, and 4.7x ± 1.6 ×compared to RM2 lib/Dfam for all analyzed species (Table 3 and Supplementary table S15). As such, our current workflow and Actiniaria-REPLib highlight the benefits of combining several tools for detection and annotation REPs. Re-annotation of unclassified TEs using TEsorter and DeepTE yielded a high level of success (Fig. 3) but with conflicting results for 265 TE entries. DANTE and TEclass2 provided consistent improvements in annotations, highlighting the effectiveness of combining protein domain-based, *k-mers* and convolutional neural networks (CNNs) in pipelines. Our strategy is effective for analyzing TEs in actinarian species, regardless whether the data are low-coverage sequencing or a high-quality genome assembly, and it enhances TE class or superfamily annotation without affecting the determination of repetitive sequences. This more precise accounting of REP sequence provides a higher resolution understanding of actiniarian genomes and will assist future studies of genomic adaptation and studies of novelties with neutral effects.

Taking into account the 38 assemblies used to construct the REPlib, we annotated 24 assemblies for the first time and re-annotated and deposited/released 14 assemblies: only two of these species are represented in Dfam and literature (*A. sola*, *N. vectensis*) [43, 44, 62, 86], one in both databases and literature (*N. vectensis*) and 12 represented in the literature (*Actinernus* sp., *Actinoscyphia* sp., *A. idsseensis*, *A. sola*, *A. tenebrosa, E. diaphana*, *E. elegans*, *M. farcimen, M. senile, P. xishaensis, S. callimorphus, T. stephensoni*) [87–97]. Of these three, we could only find data for *E. diaphana*, which released their REP as a JBrowse track [98]; the rest, as far as we could determine, presented numeric values in their results section, but did not provide access to the curated repeat data (Table 3 and Supplementary Table S15). In fact, most cnidarian REPs have not been deposited in specific repetitive content databases as Dfam and Repbase nor in specific project-based databases (e.g., Medusozoa: [51]). As such, we are unable to evaluate these annotations nor compare and reuse them if they outperformed Actiniaria-REPlib. If we compare our main results with those deposited and published, Actiniaria-REPlib identifies and classifies more

repeats than Repbase, RM2 lib, or original results from literature (Table 3 and Supplementary Table S15).

Dfam and Repbase have important differences between themselves, in addition to their annotation differences with our custom Actiniaria-specific database. Repbase includes fewer species and lower numbers of repeats but is expected to be higher quality because of its manual curation. On the other hand, Dfam is an open-access collection that offers both curated and uncurated versions and where researchers can submit and contribute their own annotations (potentially improving those already deposited in Dfam). We think that pipelines as Actiniaria-REPlib offer the benefits of each of these strategies, with the additional advantages of presenting all the details of the material and methods, allowing alternative annotation styles and potential deposition in an fully open-access database (Dfam).

Evolutionary REP trends in actiniarian genomes

In combination with the classification of REP sequences provided by Actiniaria-REPLib, our phylogeny helps contextualize differences in genomes and points to future macroevolutionary questions. Our analyses identify some intriguing differences that warrant further study. For instance, *A. sola* has a remarkably high relative amount of RC/Helitron (7.85% vs ~0.8% rest of analyzed species). This species has diverged recently from *Anthopleura elegantissima* (see McFadden et al. [99]). Comparing the REP content of *A. sola* and *A. elegantissima* could reveal whether REP expansion is linked to their speciation or a shared genomic trait. It may also indicate whether this pattern remains consistent across *A. elegantissima*'s range or evolves in isolation or response to environmental variation. We see relatively small genomes and smaller REP repertoires in *E. diaphana* and *D. lineata*, which belong to the same superfamily and which both have important ecological roles as invasive species (see Glon et al. [100]). In contrast, *Actinernus* sp., *Actinoscyphia* sp., and *P. xishaensis* have relatively larger genomes compared to other actiniarians (except *S. callimorphus*) and a higher REP proportion of ~62% (vs. 40–50% rest of species, except *E. diaphana* and *D. lineata*).

The genomes *A. alcyonoideum*, *A. arboreum*, *Actinoscyphia* sp., *R. magnifica, S. helianthus*, and *S. mertensii* have relatively higher amounts of LTR, compared to the other species (9.2–11.6% vs 2.1–8.9% rest of species). The genomes of *A. tenebrosa*, *E. elegans*, and *S. tapetum* contain relatively higher amounts of rRNA than the rest of species (> 1.3%). The inferred size of the genome is fairly consistent across the sampled species, with a few outliers: *Actinernus* sp., *Actinoscyphia* sp., *A. arboreum*, *E. quadricolor, P. xishaensis*, and *S. callimorphus* have relatively large genomes, and *S. diademon* has a relatively small genome (Fig. 1 and Table 4). Perhaps because they are inferred to be approximately twice the size of the genomes of other species, the genomes of *Actinernus* sp., *A. arboreum*, *Actinoscyphia* sp., *A. viridis*, *D. cincta*, *D. leucolena*, *E. elegans*, *P. xishaensis*, and *S. callimorphus* present a substantially higher amount of "repeats under 0.001%" (8.7–20.2% vs ~4% rest of species). Further study of the genome and REP in these organisms, in light of their phylogeny, may illuminate the historical dynamics and the role of repetitive sequences in shaping evolutionary trends.

Repeat landscapes for the repetitive sequences in each species' genome reveal the abundance of various genomic variants across levels of divergence (Fig. 2). Assuming that repeat sequence evolution is primarily driven by point mutations (which increase sequence divergence) and homogenizing amplification (which decreases intraspecific divergence), it is logical to infer that the repeat landscape for a given element reflects temporal changes in abundance. The repeat landscapes show instances of amplification of TE copies throughout the genomes, referred to as REP bursts. Across genomes, a recent REP burst within the 0–10% divergence range has been observed for DNA transposons followed by LTRs (Fig. 2). Notably, we observed a recent species-specific REP burst of RC/Helitron in the *A. sola* genome (Fig. 2), indicating a derived evolutionary condition within this genome.

Final conclusion

To our knowledge, this full-scale annotation strategy is the first effort for a cnidarian clade. This context reinforces that, even though knowledge of the REP is a growing research area with space for improvement, pipelines like Orthoptera-TElib [67] and our own present advances in several theoretical and practical fronts. Given how we have structured Actiniaria-REPlib and our strategy to reclassify assemblies, we can recognize more content and genomic positions for original datasets and an enriched comparisons with other cnidarians.

Key questions that the REP may help answer include how certain lineages have accumulated different pools of genetic elements, and how these may have been repurposed over evolutionary time for new functions or regulatory roles (including enhancing genomic plasticity). In the future, manual curation efforts in repeatome libraries and a wider phylogenetic sampling of actinarian genomes should lead to updated versions of Acinaria-REPlib. This effort should also provide motivation and a framework for developing repeat libraries for other major lineages within Cnidaria.

Material and methods

Genome of *Actinostella flosculifera* (Le Seuer, 1817)

Sample collection, DNA extraction, and sequencing

We collected one individual of *Actinostella flosculifera* from Praia do Lamberto, Saco da Ribeira, Ubatuba, São

Table 4 Annotation and comparison of 36 actiniarian genomes using Actiniaria-REPlib_v1 library in dnaPipeTE (Figure 3). Value is related to the proportion of the genome of each species. Abbreviations– LC: Low complexity; RU: Repeats under 0.001%; REP: Repeatome; Sat: Satellite; SR: Simple repeat; UNK: Unknown. Superfamilies– Actinernoidea (green), Actinioidea (pink), Actinostoloidea (yellow), Edwardsioidea (purple), and Metridioidea (blue)

Taxonomy		Genome -size (Mbp)	dnaPipeTE (%)																											
			Transposable elements															Tandem repeat sequences												
				Retrotransposons									DNA transposons																	
					LTR	non-LTR							DNA	RC/ Helitron		rRNA		Sat		SR		snRNA		LC		RU		UNK		Total REP	Non-repeat
						DIRS	PLE	LINE	SINE																					
Family	Species																													
Actinernidae	*Actinernus* sp.	1,400	1.5	0.1	6.4	0.0	0.2	0.4	3.8	0.3	25.6	0.1	0.2	0.0	0.1	0.0	0.0	0.0	20.2	3.2	62.0	38.0
Edwardsiidae	*Edwardsia elegans*	397	0.1	0.1	6.8	0.0	0.3	0.2	3.9	0.3	22.5	0.7	1.4	0.0	0.1	0.0	0.0	0.0	8.7	6.4	51.2	48.8
	Nematostella vectensis	270	0.1	0.1	5.5	0.0	0.1	0.5	3.1	0.3	30.6	0.2	0.7	0.1	0.1	0.0	0.0	0.0	2.4	2.6	46.3	53.7
	Scolanthus callimorphus	596	0.0	0.1	5.0	0.0	0.2	0.4	2.3	0.2	16.7	0.5	0.7	0.1	0.0	0.0	0.0	0.0	9.4	3.7	39.3	60.7
Actiniidae	*Anemonia viridis*	401	0.1	0.1	7.4	0.0	0.6	0.9	2.1	0.4	18.3	0.5	0.9	0.0	0.0	0.0	0.0	0.0	10.2	7.0	48.7	51.3
	Actinia mediterranea	282	0.1	0.1	6.1	0.0	0.2	0.4	2.2	0.2	19.8	0.1	0.4	0.0	0.1	0.0	0.0	0.0	4.3	3.5	37.6	62.4
	Actinia tenebrosa	240	0.3	0.1	6.8	0.0	0.2	0.4	1.6	0.2	23.9	0.0	1.7	0.0	1.0	0.0	0.0	0.0	1.0	3.8	40.9	59.1
	Actinostella flosculifera	270	0.0	0.2	7.3	0.0	0.1	0.1	1.2	0.2	27.1	0.5	0.0	0.0	0.0	0.0	0.0	0.0	2.9	6.1	45.8	54.2
	Anthopleura artemisia	342	0.1	0.2	8.1	0.0	0.1	0.2	1.6	0.3	21.6	1.3	0.0	0.0	0.0	0.0	0.0	0.0	4.5	3.2	41.3	58.7
	Anthopleura sola	290	0.2	0.2	4.1	0.0	0.0	0.2	1.1	0.4	14.6	7.9	0.1	0.0	0.0	0.0	0.0	0.0	3.2	2.5	34.6	65.4
	Bunodosoma granuliferum	352	0.0	0.1	8.4	0.0	0.2	0.3	2.1	0.7	21.6	0.3	0.2	0.0	0.0	0.0	0.0	0.0	4.6	5.2	43.7	56.3
	Condylactis gigantea	239	0.0	0.3	6.9	0.0	0.3	0.7	1.5	0.4	19.1	0.3	0.1	0.0	0.0	0.0	0.0	0.0	3.2	4.4	37.3	62.7
	Entacmaea quadricolor	430	0.0	0.1	7.8	0.0	0.5	1.8	4.2	0.7	24.6	0.2	0.5	0.0	0.0	0.0	0.0	0.0	5.6	7.0	52.8	47.2
	Urticina crassicornis	302	0.0	0.1	8.9	0.0	0.3	0.2	2.1	0.3	18.2	0.4	0.1	0.0	0.0	0.0	0.0	0.0	3.8	6.8	41.2	58.8
Actinodendridae	*Actinodendron alcyonoideum*	370	0.0	0.2	11.4	0.0	0.1	1.3	1.8	0.1	25.2	0.0	0.3	0.0	0.0	0.2	0.0	0.0	3.6	4.5	48.5	51.5
	Actinodendron arboreum	628	0.0	0.4	11.6	0.1	0.2	1.6	2.8	0.3	17.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	10.8	6.2	51.1	48.9

Table 4 (continued)

Heteractidae	***Heteractis aurora***	248	0.1	0.1	7.4	0.0	0.3	0.9	2.6	0.8	27.6	0.1	0.0	0.0	0.0	0.0	0.0	2.1	5.1	47.1	52.9
	Heteranthus verruculatus	411	0.1	0.1	5.6	0.0	0.2	0.7	1.8	0.4	14.8	0.8	0.0	0.0	0.0	0.0	0.0	7.3	5.0	36.5	63.5
	Radianthus crispa	280	0.1	0.2	8.9	0.0	0.2	1.2	1.8	0.9	21.9	0.3	0.0	0.0	0.1	0.1	0.0	3.1	5.6	44.2	55.9
	Radianthus magnifica	278	0.1	0.5	9.4	0.0	0.4	1.3	2.2	0.3	21.1	0.4	0.0	0.1	0.0	0.0	0.0	3.6	5.4	44.8	55.2
Phymanthidae	***Phymanthus crucifer***	297	0.0	0.3	6.7	0.0	0.2	0.9	1.5	0.7	20.3	0.2	0.0	0.0	0.0	0.0	0.0	4.2	5.6	40.8	59.3
	Phymanthus loligo	320	0.0	0.1	6.5	0.0	0.1	0.5	1.2	0.9	23.1	0.4	0.0	0.0	0.0	0.0	0.0	4.0	3.2	40.2	59.8
Stichodactylidae	***Stichodactyla helianthus***	230	0.0	0.1	9.5	0.0	0.3	1.4	2.3	0.5	21.3	0.3	0.0	0.0	0.0	0.0	0.0	2.4	4.9	42.9	57.1
	Stichodactyla mertensii	295	0.0	0.1	9.3	0.0	0.4	1.1	3.6	0.5	21.9	0.2	0.0	0.0	0.0	0.0	0.0	3.5	5.5	46.1	53.9
	Stichodactyla tapetum	333	0.0	0.1	7.2	0.0	0.5	1.4	2.0	0.5	19.5	0.6	1.4	0.0	0.1	0.0	0.0	4.8	5.4	43.6	56.4
	***Thalassianthus aster* (=*Stichodactyla* sp.)**	262	0.0	0.4	5.6	0.0	0.5	1.1	2.5	0.6	25.5	0.4	0.0	0.0	6.3	0.0	0.0	2.6	4.5	49.9	50.1
Actinostolidae	***Stomphia didemon***	160	0.0	0.1	5.4	0.0	0.4	0.2	0.8	0.2	24.9	0.3	0.0	0.0	0.0	0.0	0.0	1.3	4.3	37.9	62.1
Actinoscyphiidae	***Actinoscyphia* sp.**	522	0.2	0.2	10.8	0.0	0.5	0.5	7.7	0.2	23.4	0.7	0.0	0.0	0.0	0.0	0.0	11.7	5.9	61.7	38.3
Aiptasiidae	***Aiptasiogeton hyalinus***	249	0.0	0.1	4.7	0.0	0.3	0.6	1.4	0.0	14.8	0.4	0.3	0.0	0.0	0.0	0.0	4.6	5.6	32.8	67.2
	Exaiptasia diaphana	260	0.0	0.0	2.1	0.0	0.0	0.2	0.3	0.2	6.4	0.1	0.1	0.0	0.0	0.0	0.0	2.6	2.0	13.9	86.1
Diadumenidae	***Diadumene cincta***	366	0.0	0.1	7.7	0.0	0.3	0.7	3.7	0.2	18.4	1.9	0.3	0.0	0.0	0.0	0.0	9.1	4.6	46.9	53.1
	Diadumene leucolena	360	0.0	6.4	6.4	0.0	0.3	0.4	3.1	0.2	20.2	1.6	0.3	0.0	0.0	0.0	0.0	9.0	2.5	50.4	49.6
	Diadumene lineata	313	0.0	0.1	2.2	0.0	0.1	0.2	0.8	0.3	10.3	0.3	0.1	0.0	0.0	0.0	0.0	4.3	1.2	19.8	80.2
Hormathiidae	***Paraphelliactis xishaensis***	543	0.1	0.2	6.1	0.0	0.3	0.3	4.5	0.2	29.0	0.4	0.6	0.0	0.0	0.0	0.0	10.1	5.1	56.9	43.1
Metridiidae	***Metridium farcimen***	340	0.1	0.1	5.2	0.0	0.1	0.7	2.8	0.3	27.2	0.4	0.5	0.0	0.1	0.0	0.0	4.2	3.7	45.3	54.7
	Metridium senile	390	0.0	0.1	6.0	0.0	0.2	0.5	6.0	0.2	28.2	0.2	0.7	0.1	0.1	0.1	0.0	4.8	2.1	49.0	51.0

Paulo (USP, 23°30′04.6"S, 45°07′09.1"W), on July 8, 2022. This animal was kept in an aquarium at the Laboratory of Evolution and Aquatic Diversity (LEDALab), São Paulo State University (UNESP-Bauru), fed *Artemia* sp. and bivalves two to three times per week over several months. Feeding was stopped three days prior to DNA extraction to avoid exogenous DNA.

We isolated total genomic DNA of *A. flosculifera* from a 200 mg piece of fresh (live) tissue using the QIAamp® DNA Mini Kit (QIAGEN) (RRID:SCR_008539). Library preparation, sequencing, and raw data control were done by IntegraGen SA (Evry, France) according to supplier recommendations based on a PCR-free strategy. Briefly, they prepared libraries using NEBNext Ultra II DNA Library Prep Kits (NEB #E7103). They quantified double-strand gDNA and used a sonication method to fragment approximately 520 ng of high-molecular-weight gDNA into ~400 bp fragments. They ligated paired-end adaptor oligonucleotides (xGenTM TS-LT Adapter Duplexes (IDT #1,077,681)) and re-paired them. The tailed fragments were purified for direct sequencing without a PCR step. They sequenced the libraries on an Illumina NovaSeq platform, generating ~294 million 2× 150 bp paired-end reads. Finally, image analysis and base calling were performed using Illumina Real Time Analysis (RTA) Pipeline version 3.4.4 with default parameters.

Sequencing data pre-processing (Fig. 3, Stage I–I')

We applied the "LEDAlabShortReadDecontamination" [101] pipeline for processing Illumina sequencing reads as follows: we trimmed the FASTQ files with fastp (RRID:SCR_016962) v0.23.4 [102] and we concatenated the two unpaired FASTQ files using Contig Annotation Tool (CAT) v5.3 [103]; we assessed read quality before and after processing with FastQC (RRID:SCR_014583) v0.12.1 [104], MultiQC (RRID:SCR_014982) v1.20 [105], and SeqKit (RRID:SCR_018926) v2.8.0 [106]; we used ALLPATHS-LG (RRID:SCR_010742) v.52488 ErrorCorrectReads.pl script [107] to apply error correction to reads; we used Kraken2 (RRID:SCR_005484) v2.1.3 [108] to create and build a database (DB_library), and to remove exogenous DNA from the FASTQ files (see Supplementary material S2 and Supplementary Table S16); finally, we assembled the *A. flosculifera* mitogenome with GetOrganelle v1.7.7.0 [109] using the *Actinia tenebrosa* Farquhar, 1898 mitogenome as 'seed' (available in NCBI with accession number NC_044902.1), and then removed the *A. flosculifera* mitogenome reads of the original paired and unpaired reads using FastqSifter (RRID:SCR_017200) [110]. Same basic protocol was used to isolate original reads and assemble the mitochondrial genome for several species included in this study (Table 2) to prepare for subsequent mitochondrial DNA annotation (see below).

Mitogenome annotation

We annotated the 36 assembled mitogenomes using MITOS2 [111] with the mitochondrial genetic code of "Mold, Protozoa, and Coelenteral", and the reference data"RefSeq89 Metazoa", with default parameters to predict protein-coding genes (PCGs), tRNAs, and rRNAs genes. We compared the control region of the mitochondrial genome, designated as blank region, with mitochondrial genomes of reference species within Actiniaria, including *Actinia tenebrosa* (GenBank NC_044902.1), *Exaiptasia diaphana* (Rapp, 1829) (GenBank NC_056771.1), and *Nematostella vectensis* (GenBank NC_008164.1). We determined the starting position and orientation of the mitochondrial assembly sequence using Geneious Prime (RRID:SCR_010519) v2024 [112]. Finally, we deposited the complete, annotated, mitochondrial DNA sequence of the 27 species that were not included at NCBI database under the accession number in Table 2.

Phylogenetic reconstruction

We used the 13 protein coding genes (ND1–6, COX1–3, CYTB, and ATP6 + 8) and 2 rRNA (12S and 16S) of each of the 36 assembled and annotated mitogenomes (Supplementary Material S1) for a phylogenetic reconstruction of Actiniaria. We aligned each gene with MAFFT (RRID:SCR_011811) v7.53 using the L-INS-I algorithm and the "--maxiterate 1000" option [113]. We concatenated the aligned genes in a matrix using SequenceMatrix v1.8 [114] (Supplementary Material S1). Selection of the best partition strategy and evolutionary model (see details in Supplementary Material S1) was based on the best Bayesian Information Criterion (BIC) score using ModelFinder and PartitionFinder [115] as implemented in IQ-Tree2 (RRID:SCR_017254) [116] (Supplementary Material S1); we used this same software for Maximum likelihood (ML) phylogenetic inference and branch support. For these analyses, we applied (i) nonparametric approaches SH-like approximate likelihood ratio test (SH-aLRT; 1000 replicates) and ultrafast bootstrap (UFBoot2, 1000 replicates) [117]; (ii) parametric approximate likelihood ratio test (aLRT) and approximate Bayes tests (aBAYES), 1000 replicates for both cases [118, 119] (Supplementary Material S1). We edited and visualized the resulting tree using TreeGraph v2.15 [120].

Genome assembly (Fig. 3, Stage II)

We used the "RyanLabShortReadAssembly" pipeline [121], as a guide for assembling the Illumina sequencing reads from the previous stage (nuclear, "decontaminated" reads-only): i) we calculated the *k-mer* counts (sizes 21, 25, 31, 45, 63, 81 and 99) occurrence of the DNA in FASTQ files using Jellyfish (RRID:SCR_005491)

Table 5 Statistics for the genome assembly of *Actinostella flosculifera*

NCBI Taxa ID	**3,034,631**
No. of sequences	62,998
Estimated genome size (bp)	269,371,768
Longest sequence (bp)	12,947,013
N50 scaffold (bp)	13,099
BUSCO (% complete)	70.55
BUSCO (% complete + partial)	91.61
GC content (%)	38.8
Assembly accession	JBLZGT010000000
NCBI raw read accession	SRR31542901
Specimen Voucher ID	Aflosc_v1

v2.3.1 [122]; ii) we parsed the resulting *k-mer* count histograms in GenomeScope (RRID:SCR_017014) [123] so that we could visualize their distribution; iii) we generated nine assemblies using Platanus (RRID:SCR_01553) v1.2.4 (plat.pl) [124] with *k-mer* sizes of 21, 25, 31, 45, 59, 63, 73, 81, and 99 and we used Redundans v2.01 [125] to selectively remove alternative heterozygous contigs by running "redundans.py" in each assembled genome with the different *k-mer*s; iv) we choose the best *k-mer* of nine assemblies based on N50 and conserved orthologs using BUSCO (RRID:SCR_015008) v5 [126] through the online platform gVolante [127]; v) we used the remaining assemblies (e.g., the sub-optimal assemblies) to construct artificial mate-pair libraries of 3 insert sizes (2000, 5000, and 10,000) with Matemaker (RRID:SCR_017199) v1.2 [128]; vi) we used the artificial mate-pair libraries to scaffold the optimal assembly (generated using Platanus of the best *k-mer*) with SSPACE Standard (RRID:SCR_005056) v3.0 [129]; vii) we removed sequences shorter than 200 bp in the scaffold using remove_short_and_sort from the RyanLabShortReadAssembly pipeline; viii) finally, we use this assembly to produce reference-guided scaffolds using RagTag v2.1.0 [130] with the scaffold-level assembly from a confamilial species, *Anthopleura sola* (GCA_023349385.1), as a reference. Improvements on this last assembly step was assessed with N50 and BUSCO metrics as well.

De novo construction of the Actiniaria-REPlib library (Fig. 3, Stage III)

We built the Actiniaria-specific REP library (named Actiniaria-REPlib) de novo based on 38 assemblies (Table 2) following the general strategy developed for Orthoptera-TElib pipeline (see Liu et al. [67] for details). We analyzed 37 actiniarian genomic datasets available at NCBI-Dataset ([60]; accessed 11.15.2024) and the newly generated assembly of *A. flosculifera* (Table 5). To predict TEs, we used RepeatModeler2 (RRID:SCR_015027) [131] for each of the 38 genomes using Dfam v3.8 partition 0 (dfam38_full.0.h5) [44]. We merged the REP libraries generated for each of the species into one initial REP library (RM2 lib) using CAT v6.0.1 [103] (version Actiniaria-REPlib_A). From this, we removed redundant sequences using CD-hit (RRID:SCR_007105) v4.8.1 [132] applying the 80–80–80 rule [17], saving this as Actiniaria-REPlib_B). We separated unknown sequences from Actiniaria-REPlib_B library with Seqtk (RRID:SCR_018927) v1.4 [133] and re-annotated them with TEsorter v1.4.6 [134] and DeepTE [135]. Then, we used Domain Based Annotation of Transposable Elements (DANTE v0.9.1) [136] (-D Metazoa_v3.1) and TEclass2 [137] to re-annotate the conflicting sequences based on the mismatch annotations between TEsorter and DeepTE. We merged the Actiniaria-REPlib_B_known library, DeepTE non-conflicting annotation library, and re-annotated sequences by DANTE + TEclass2. This is the first version of the REP library for the Actiniaria clade called Actiniaria-REPlib (or Actiniaria-REPlib_v1).

Annotation and quantification of the REP content (Fig. 3, Stage IV)

We evaluated and compared the annotation efficiency of our aforementioned three REP libraries (RM2 lib, Actiniaria-REPlib_A, Actiniaria-REPlib_B) for the original, full dataset of 38 assembled actiniarian genomes. Also, we compared our new database (Actiniaria-REPlib) to Repbase (RRID:SCR_021169) v29.03 specific to *Nematostella vectensis* and RM2 lib using RepeatMasker (RRID:SCR_012954) v4.1.6 [138].

We further applied the dnaPiPeTE v1.3.1 pipeline [68] to classify and quantify repeats in 36 actiniarian genomes using Illumina sequencing reads for comparative analysis (several genomes assemblies did not have Illumina reads available; Table 2). We pre-processed the Illumina sequencing reads of the 36 actiniarian species (Table 2), following the pre-processing methods used for *A. flosculifera* in mitogenome assembly, trimming, error correction, and exogenous DNA removal (see above; Fig. 3, Stage I–I'). We used 0.25 ×genome coverage Illumina sequencing reads, Actiniaria-REPlib and genome-size as input in dnaPiPeTE (Fig. 3, Stage IV). The genome size was determined using the value obtained from the NCBI assemblies. We used dnaPT_charts.sh [139, 140] to plot the relative proportions of each assembled repeat. To generate repeat landscapes, we plotted histograms with dnaPT_landscapes.sh [139, 140] that represent the BLASTN divergence measured between each TE copy in each genome and read and their consensus assembled repeats [68, 141].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11591-0.

Supplementary Material S1. The results of the phylogenomic analysis of Actiniaria.

Supplementary Material S2. The outputs of the *Actinostella flosculifera* genome assembly results.

Supplementary Table S1. Cnidarian taxa included in Dfam v3.8 partition 0 (dfam38_full.0.h5) and Repbase v.29.03, and curated and uncurated annotations of REPs. Species present in both databases are highlighted in bold. Supplementary Table S2. Sequencing data pre-processing and assembling workflow of *Actinostella flosculifera*. Red numbers are PAIRED data and green numbers are UNPAIRED data. Supplementary Table S3. Gene structure of *Actinostella flosculifera*. Supplementary Table S4. Sequencing data pre-processing workflow of the 35 genomes, not including *Actinostella flosculifera* genome. ND: No data. Supplementary Table S5. Gene structure of the new mitogenome of 26 species that were not included in the NCBI database. Supplementary Table S6. The 38 actiniaria species and REP libraries built by RepeatModeler2. Outputs of each genome are in the figshare repository (dx.doi.org/10.6084/m9.figshare.27011698). Supplementary Table S7. Output at the non-redundant library construction stage of known sequences (N=13,601). Supplementary Table S8. Re-annotated results using 66,299 unknown through DeepTE. Supplementary Table S9. Re-annotated results using 66,299 unknown through TEsorter. Supplementary Table S10. The 346 REP entries were annotated by the DeepTE and TEsorter packages. Blue highlighting of conflicting REP annotations (*n*=265). Supplementary Table S11. TEclass2 re-annotation results of 256 conflicting REP entries. Supplementary Table S12. DANTE re-annotation results of 256 conflicting REP entries. Supplementary Table S13. The result of the final annotation of 256 conflicting REP entries. Supplementary Table S14. Classification and annotation of REP in Actiniaria-REPlib_v1 and comparison with the classification format of Repbase and Dfam. Nr sequences- known: 74,643; unknown: 5,260. NA: No data. Supplementary Table S15. Efficiency in the annotation of three libraries of REPs (Repbase, the library built by RepeatModeler2 of each genome (RM2lib), and Actiniaria-REPlib_v1) in 38 actinarian genomes and results of previous studies (in bold) (Figure 4). Supplementary Table S16. Taxa added to build DB_library by Kraken2 (together with human, bacteria, viral, uniVec, archaea, plasmid libraries available in the NCBI database, see protocol in Supplementary Material S1).

Acknowledgements
The computational resources provided by Ohio Supercomputer Center (OSC) [142] at The Ohio State University are gratefully acknowledged. We thank Dr. Michael Broe for his help in solving the bioinformatics problems at the OSC.

Authors' contributions
J.A.D.F and S.N.S. collected the samples; J.A.D.F and M.M.M conceived the idea; J.A.D.F conducted the bioinformatic work; J.A.D.F. and M.M.M. analyzed the data and led the writing with the support of O.M.P.G, E.R.C, M.M.M., J.F.R., MD, and S.N.S.; J.A.D.F. and M.M.M. accept full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish; all authors reviewed and approved the final version of the manuscript.

Funding
This study was supported by São Paulo Research Foundation (FAPESP) [Proc. n. 2019/03552-0, 2020/16589-7, 2022/09430-7, 2022/16193-1 and 2023/10683-0]. SNS was supported by the National Council of Scientific and Technological Development (CNPq -Research Productivity Scholarship) grant number 304267/2022-8. MMM was funded by FAPESP 2016/04560-9 and PROPe-UNESP grant number 4390. OMPG was supported by the Swedish Research Council Vetenskapsrådet (grant number 2020-03866).

Data availability
All data supporting the findings of this study is available on Figshare under the identifier https://doi.org/10.6084/m9.figshare.27011698 (ref. [140]). Final Actiniaria REP library with alternative nomenclatures (Dfam, Repbase and Actiniaria-REPlib_v1) is shared in Supplementary Table S14. We deposited the complete, annotated, mitochondrial DNA sequence of the 27 species to the NCBI database under the accession numbers that are included in Table 2. Actinostella flosculifera raw data: SRA Genbank SRR31542901. Bioinformatic codes are available at https://github.com/jefferalexdurfue/LEDAlabShortReadDecontamination (ref. [101]).

Declarations

Ethics approval and consent to participate
All applicable international, national, and/or institutional guidelines for the care and use of animals were followed by the authors. All necessary permits for sampling and observational field studies have been obtained by the authors from the competent authorities and are mentioned in the acknowledgements, if applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 26 December 2024 Accepted: 9 April 2025
Published online: 13 May 2025

References

1. Gregory TR. The evolution of the genome. Elsevier. 2005. https://doi.org/10.1016/B978-0-12-301463-4.X5000-1.
2. Graur D, Sater AK, Cooper TF. Molecular and genome evolution. Massachusetts, USA: Sinauer Associates, Incorporated; 2016.
3. Jeffery NW, Jardine CB, Gregory TR. A first exploration of genome size diversity in sponges. Genome. 2013;56:451–6. https://doi.org/10.1139/gen-2012-0122.
4. Doležel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry. Nat Protoc. 2007;2233–2244. https://doi.org/10.1038/nprot.2007.310
5. Pflug JM, Holmes VR, Burrus C, Johnston JS, Maddison DR. Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). G3: Genes, Genomes, Genetics. 2020;10:3047–60. https://doi.org/10.1534/g3.120.401028.
6. Graur D, Zheng Y, Azevedo RB. An evolutionary classification of genomic function. GBE. 2015;7:642–5. https://doi.org/10.1093/gbe/evv021.
7. Siefert JL. Defining the Mobilome. In: Gogarten, M.B., Gogarten, J.P., Olendzenski, L.C. (eds) Horizontal Gene Transfer. Methods Mol Biol. 2009;532. Humana Press. https://doi.org/10.1007/978-1-60327-853-9_2
8. Elliott TA, Gregory TR. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos Trans R Soc Lond B Biol Sci. 2015;370:20140331. https://doi.org/10.1098/rstb.2014.0331.
9. Dunn CW, Ryan JF. The evolution of animal genomes. Curr Opin Genet Dev. 2015;35:25–32. https://doi.org/10.1016/j.gde.2015.08.006.
10. Maumus F, Quesneville H. Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS ONE. 2014;9: e94101. https://doi.org/10.1371/journal.pone.0094101.
11. Biscotti MA, Olmo E, Heslop-Harrison JS. Repetitive DNA in eukaryotic genomes. Chromosome Res. 2015;23:415–20. https://doi.org/10.1007/s10577-015-9499-z.
12. Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. The struggle for life of the genome's selfish architects. Biol Direct. 2011;6:19. https://doi.org/10.1186/1745-6150-6-19.
13. Orgel L, Crick F. Selfish DNA: the ultimate parasite. Nature. 1980;284:604–7. https://doi.org/10.1038/284604a0.
14. Kidwell, MG. Transposable Elements. In: The evolution of the genome. Academic Press 2005;165–221. https://doi.org/10.1016/B978-012301463-4/50005-X
15. Kazazian HH. Mobile Elements: Drivers of Genome Evolution. Science. 2004;303:1626–32. https://doi.org/10.1126/science.1089670.

16. Suh A. Genome size evolution: small transposons with large consequences. Curr Biolog. 2019;29:R241–3. https://doi.org/10.1016/j.cub.2019.02.032.
17. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Schulman AH. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82. https://doi.org/10.1038/nrg2165.
18. Rubin GM, Spradling AC. Genetic transformation of *Drosophila* with transposable element vectors. Science. 1982;218:348–53. https://doi.org/10.1126/science.6289436.
19. Galbraith JD, Hayward A. The influence of transposable elements on animal colouration. TiG. 2023;39:624–38. https://doi.org/10.1016/j.tig.2023.04.005.
20. Liu P, Cuerda-Gil D, Shahid S, Slotkin RK. The epigenetic control of the transposable element life cycle in plant genomes and beyond. Annu Rev Genet. 2022;56:63–87. https://doi.org/10.1146/annurev-genet-072920-015534.
21. Senft AD, Macfarlan TS. Transposable elements shape the evolution of mammalian development. Nat Rev Genet. 2021;22:691–711. https://doi.org/10.1038/s41576-021-00385-1.
22. Bennetzen JL. Transposable element contributions to plant gene and genome evolution. Plant Mol Biol. 2000;42:251–69. https://doi.org/10.1023/A:1006344508454.
23. Hewitt GM. Population cytogenetics. Curr Opin Genet Dev. 1992;2:844–9. https://doi.org/10.1016/S0959-437X(05)80105-4.
24. Montgomery EA, Huang SM, Langley CH, Judd BH. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. Genetics. 1991;129:1085–98. https://doi.org/10.1093/genetics/129.4.1085.
25. Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. 1994;371:215–20. https://doi.org/10.1038/371215a0.
26. Khost DE, Eickbush DG, Larracuente AM. Single-molecule sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila melanogaster*. Genome res. 2017;27:709–21. https://doi.org/10.1101/gr.213512.116.
27. Fingerhut JM, Yamashita YM. The regulation and potential functions of intronic satellite DNA. Semin Cell Dev Biol. 2022;128:69–77. https://doi.org/10.1016/j.semcdb.2022.04.010. Academic Press.
28. Dover G. Molecular drive. TIG. 2002;18:587–9. https://doi.org/10.1016/S0168-9525(02)02789-0.
29. Ugarković Đ, Plohl M. Variation in satellite DNA profiles—causes and effects. EMBO J. 2002;21:5955–9. https://doi.org/10.1093/emboj/cdf612.
30. Dover G. Molecular drive: a cohesive mode of species evolution. Nature. 1982;299:111–7. https://doi.org/10.1038/299111a0.
31. Lower SS, McGurk MP, Clark AG, Barbash DA. Satellite DNA evolution: old ideas, new approaches. Curr Opin Genet Dev. 2018;49:70–8. https://doi.org/10.1016/j.gde.2018.03.003.
32. Plohl M, Luchetti A, Meštrović N, Mantovani B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. Gene. 2008;409:72–82. https://doi.org/10.1016/j.gene.2007.11.013.
33. Walsh JM. Persistence of Tandem Arrays: Implications for Satellite and Simple-Sequence DNAs. Genetics. 1987;115:553–67. https://doi.org/10.1093/genetics/115.3.553.
34. Smith GP. Evolution of Repeated DNA Sequences by Unequal Crossover. Science. 1976;191:528–35. https://doi.org/10.1126/science.1251186.
35. Palacios-Gimenez OM, Milani D, Song H, Marti DA, López-León MD, Ruiz-Ruano FJ, Camacho JPM, Cabral-de-Mello DC. Eight million years of satellite DNA evolution in grasshoppers of the genus *Schistocerca* illuminate the ins and outs of the library hypothesis. GBE. 2020;12:88–102. https://doi.org/10.1093/gbe/evaa018.
36. Palacios-Gimenez OM, Koelman J, Palmada-Flores M, Bradford TM, Jones KK, Cooper SJB, Kawakami T, Suh A. Comparative analysis of morabine grasshopper genomes reveals highly abundant transposable elements and rapidly proliferating satellite DNA repeats. BMC Biol. 2020;18:199. https://doi.org/10.1186/s12915-020-00925-x.
37. Plohl M, Petrović V, Luchetti A, Ricci A, Šatović E, Passamonti M, Mantovani B. Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. Heredity. 2010;104:543–51. https://doi.org/10.1038/hdy.2009.141.
38. Petraccioli A, Odierna G, Capriglione T, Barucca M, Forconi M, Olmo E, Assunta BM. A novel satellite DNA isolated in *Pecten jacobaeus* shows high sequence similarity among molluscs. Mol Genet Genomics. 2015;290:1717–25. https://doi.org/10.1007/s00438-015-1036-4.
39. Chaves R, Ferreira D, Mendes-da-Silva A, Meles S, Adega F. FA-SAT is an old satellite DNA frozen in several Bilateria genomes. GBE. 2017;9:3073–87. https://doi.org/10.1093/gbe/evx212.
40. Lorite P, et al. Concerted evolution, a slow process for ant satellite DNA: study of the satellite DNA in the *Aphaenogaster* genus (Hymenoptera, Formicidae). Org Divers Evol. 2017;17:595–606. https://doi.org/10.1007/s13127-017-0333-7.
41. Escudeiro A, Adega F, Robinson TJ, Heslop-Harrison JS, Chaves R. Conservation, divergence, and functions of centromeric satellite DNA families in the Bovidae. GBE. 2019;11:1152–65. https://doi.org/10.1093/gbe/evz061.
42. Fry K, Salser W. Nucleotide sequences of HS-α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. Cell. 1977;12:1069–84. https://doi.org/10.1016/0092-8674(77)90170-2.
43. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462–7. https://doi.org/10.1159/000084979.
44. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA. 2021;12:1–14. https://doi.org/10.1186/s13100-020-00230-y.
45. RepeatMasker. https://www.repeatmasker.org. Accessed 10 April 2024.
46. Salser W, Bowen S, Browne D, El-Adli F, Fedoroff N, Fry K, Whitcome P. Investigation of the organization of mammalian chromosomes at the DNA sequence level. In Federation proceedings. 1976;35:23–35.
47. WoRMS. World Register of Marine Species. https://www.marinespecies.org. Accessed 23 July 2024. https://doi.org/10.14284/170
48. Adachi K, Miyake H, Kuramochi T, Mizusawa K, Okumura SI. Genome size distribution in phylum Cnidaria. Fisheries Sci. 2017;83:107–12. https://doi.org/10.1007/s12562-016-1050-4.
49. Zhang X, Jacobs DA. Broad survey of gene body and repeat methylation in Cnidaria reveals a complex evolutionary history. GBE. 2022;14:evab284. https://doi.org/10.1093/gbe/evab284.
50. Ying H, Hayward DC, Klimovich A, Bosch TC, Baldassarre L, Neeman T, Forêt S, Huttley G, Reitzel AM, Fraune S, Ball EE, Miller DJ. The role of DNA methylation in genome defense in Cnidaria and other invertebrates. Mol Biol Evol. 2022;39:msac018. https://doi.org/10.1093/molbev/msac018.
51. Santander MD, Maronna MM, Ryan JF, Andrade SC. The state of Medusozoa genomics: current evidence and future challenges. Gigascience. 2022;11:giac036. https://doi.org/10.1093/gigascience/giac036.
52. Kon-Nanjo K, Kon T, Yu TC-TK, Rodriguez-Terrones D, Falcon F, Martínez DE, Steele RE, Tanaka EM, Holstein TW, Simakov O. The dynamic genomes of Hydra and the anciently active repeat complement of animal chromosomes. bioRxiv 2024. https://doi.org/10.1101/2024.03.13.584568
53. Ahuja N, Cao X, Schultz DT, Picciani N, Lord A, Shao S, Jia K, Burdick DR, Haddock SHD, Li Y, Dunn CW. Giants among Cnidaria: Large nuclear genomes and rearranged mitochondrial genomes in siphonophores. Genome Biol Evol. 2024;6:evae048. https://doi.org/10.1093/gbe/evae048.
54. Alama-Bermejo G, Holzer AS. Advances and discoveries in myxozoan genomics. Trends Parasitol. 2021;37:552–68. https://doi.org/10.1016/j.pt.2021.01.010.
55. Guo Q, Atkinson SD, Xiao B, Zhai Y, Bartholomew JL, Gu Z. A myxozoan genome reveals mosaic evolution in a parasitic cnidarian. BMC Biol. 2022;20:51. https://doi.org/10.1186/s12915-022-01249-8.
56. Neverov AM, Panchin AY, Mikhailov KV, Batueva MD, Aleoshin VV, Panchin YV. Apoptotic gene loss in Cnidaria is associated with transition to parasitism. Sci Rep. 2023;13:8015. https://doi.org/10.1038/s41598-023-34248-y.
57. GENOMESIZE. Animal Genome Size Database. https://www.genomesize.com/. Accessed 10 October 2024.
58. The ACC. The Animal Chromosome Count database. https://cromanpa94.github.io/ACC/. Accessed 10 October 2024.
59. GoaT. Genomes on a Tree. https://goat.genomehubs.org/. Accessed 06 November 2024. https://doi.org/10.12688/wellcomeopenres.18658.1

60. O'Leary NA, Cox E, Holmes JB, Anderson WR, Falk R, Hem V, Tsuchiya MTN, Schuler GD, Zhang X, Torcivia J, Ketter A, Breen L, Cothran J, Bajwa H, Tinne J, Meric PA, Hlavina W, Schneider VA. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. Sci Data. 2024;11:732. https://doi.org/10.1038/s41597-024-03571-y.
61. Noel B, Denoeud F, Rouan A, et al. Pervasive tandem duplications and convergent evolution shape coral genomes. Genome Biol. 2023;24:123. https://doi.org/10.1186/s13059-023-02960-7.
62. Zimmermann B, Montenegro JD, Robb SM, Fropf WJ, Weilguny L, He S, Chen S, Lovegrove-Walsh J, Hill EM, Chen CY, Ragkousi K, Praher D, Fredman D, Schultz D, Moran Y, Simakov O, Genikhovich G, Gibson MC, Technau U. Topological structures and syntenic conservation in sea anemone genomes. Nat Commun. 2023;14:8270. https://doi.org/10.1038/s41467-023-44080-7.
63. He, C, Han T, Huang W, Wang B, Liao X, Chen J, Lu Z. Deciphering omics atlases to aid stony corals in response to global change. PREPRINT (Version 1) available at Research Square 2024. https://doi.org/10.21203/rs.3.rs-4037544/v1
64. Cowen LJ, Putnam HM. Bioinformatics of corals: investigating heterogeneous omics data from coral holobionts for insight into reef health and resilience. Annu Rev Biomed Data Sci. 2022;5:205–31. https://doi.org/10.1146/annurev-biodatasci-122120-030732.
65. Ying H, Cooke I, Sprungala S, Wang W, Hayward DC, Tang Y, Huttley G, Ball EE, Forêt S, Miller DJ. Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. Genome Biol. 2018;19:1–24. https://doi.org/10.1186/s13059-018-1552-8.
66. Fourreau CJL, Kise H, Santander MD, Pirro S, Maronna MM, Poliseno A, Santos MEA, Reimer JD. Genome sizes and repeatome evolution in zoantharians (Cnidaria: Hexacorallia: Zoantharia). PeerJ. 2023;11:e16188. https://doi.org/10.7717/peerj.16188.
67. Liu X, Zhao L, Majid M, Huand Y. Orthoptera-TElib: a library of Orthoptera transposable elements for TE annotation. Mob DNA. 2024;15:5. https://doi.org/10.1186/s13100-024-00316-x.
68. Goubert C. Assembly-free detection and quantification of transposable elements with dnaPipeTE. In: Branco, M.R., de Mendoza Soler, A. (eds) Transposable Elements. Methods Mol Biol 2023, vol 2607. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-2883-6_2
69. Rodríguez E, Barbeitos MS, Brugler MR, Crowley LM, Grajales A, Gusmão L, Häussermann V, Reft A, Daly M. Hidden among sea anemones: the first comprehensive phylogenetic reconstruction of the order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) reveals a novel group of Hexacorals. PLoS ONE. 2014;9: e96998. https://doi.org/10.1371/journal.pone.0096998.
70. Grajales A, Rodríguez E. Elucidating the evolutionary relationships of the Aiptasiidae, a widespread cnidarian–dinoflagellate model system (Cnidaria: Anthozoa: Actiniaria: Metridioidea). Mol Phylogenet Evol. 2016;94:252–63. https://doi.org/10.1016/j.ympev.2015.09.004.
71. Gusmão LC, Grajales A, Rodríguez E. Sea anemones through X-rays: visualization of two species of *Diadumene* (Cnidaria, Actiniaria) using micro-CT. Am Mus Novit. 2018;2018:1–47. https://doi.org/10.1206/3907.1.
72. Sanamyan NP, Sanamyan KE, Galkin SV, Ivin VV, Bocharova ES. Deep water Actiniaria (Cnidaria: Anthozoa) Sicyonis, Ophiodiscus and Tealidium: re-evaluation of Actinostolidae and related families. IZ. 2021;18:385–449. https://doi.org/10.15298/invertzool.18.4.01
73. Barragán Y, Rodríguez E, Chiodo T, Gusmão LC, Sánchez C, Lauretta D. Revision of the genus *Actinostella* (Cnidaria: Actiniaria: Actinioidea) from tropical and subtropical western Atlantic and eastern Pacific: redescriptions and synonymies. Am Mus Novit. 2024;2024:1–48.
74. Durán-Fuentes, JA, González-Muñoz, R, Daly, M, Stampar, SN. *Antholoba fabiani* sp. nov. (Actiniaria, Metridioidea, Antholobidae fam. nov.), a new species and family of sea anemone of the southwestern Atlantic, Brazil. *Mar Biodivers* 2024;54:40. https://doi.org/10.1007/s12526-024-01433-9
75. Vassallo-Avalos A, González-Muñoz R, Morrone JJ, Acuña FH, Durán-Fuentes JA, Stampar SN, Rivas G. A new species of *Anthopleura* (Cnidaria: Anthozoa: Actiniaria) from the Mexican Pacific. Mar Biodivers. 2024;54:70. https://doi.org/10.1007/s12526-024-01464-2.
76. McFadden CS, Quattrini AM, Brugler MR, Cowman PF, Dueñas LF, Kitahara MV, Paz-García DA, Reimer JD, Rodríguez E. Phylogenomics, origin, and diversification of Anthozoans (Phylum Cnidaria). Syst Biol. 2021;70:635–47. https://doi.org/10.1093/sysbio/syaa103.
77. Benedict C, Delgado A, Pen I, Vaga C, Daly M, Quattrini AM. Sea anemone (Anthozoa, Actiniaria) diversity in Mo'orea (French Polynesia). *Mol Phylogenet Evol* 2024;108118. https://doi.org/10.1016/j.ympev.2024.108118
78. Quattrini AM, Snyder KE, Purow-Ruderman R, Seiblitz IG, Hoang J, Floerke N, McFadden CS. Mito-nuclear discordance within Anthozoa, with notes on unique properties of their mitochondrial genomes. Sci Rep. 2023;13:7443. https://doi.org/10.1038/s41598-023-34059-1.
79. INSDC. International Nucleotide Sequence Database Collaboration. http://www.insdc.org/. Accessed 10 October 2024.
80. EGAG. Eukaryotic Genome Annotation Guide. https://www.ncbi.nlm.nih.gov/genbank/eukaryotic_genome_submission_annotation/. Accessed 10 October 2024.
81. Brown T, Collier KA, Cruz F, Gkanogiannis A, Joye-Dind S, Nevers Y, Saenko S, Alioto T, Bretaudeau A, Charleston M, Doan PD, Hahn C, Harrop TWR., Herron KE, Kebaso F, Libouban R, Mansueto L, Manu S, Oba A, Swarbreck D, Syme A, Zanarello F, Aury J-M, Gómez-Garrido J, Dennis AB. Genome annotation and other post-assembly workflows for the Tree of Life. 2024. BioHackrXiv Preprints https://doi.org/10.37044/osf.io/fy49g
82. Goubert C, Craig RJ, Bilat AF, Peona V, Vogan AA, Protasio AV. A beginner's guide to manual curation of transposable elements. Mob DNA. 2022;13:7. https://doi.org/10.1186/s13100-021-00259-7.
83. Peona V, et al. Teaching transposon classification as a means to crowd source the curation of repeat annotation–a tardigrade perspective. Mob DNA. 2024;15:10. https://doi.org/10.1186/s13100-024-00319-8.
84. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Wipat A. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26:541–7. https://doi.org/10.1038/nbt1360.
85. Yilmaz P, Kottmann R, Field D, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol. 2011;29:415–20. https://doi.org/10.1038/nbt.1823.
86. Cornwell BH, Beraut E, Fairbairn C, Nguyen O, Marimuthu MP, Escalona M, Toffelmier E. Reference genome assembly of the sunburst anemone. Anthopleura sola J Hered. 2022;113:699–705. https://doi.org/10.1093/jhered/esac050.
87. Law STS, Yu Y, Nong W, So WL, Li Y, Swale T, Hui JHL. The genome of the deep-sea anemone *Actinernus* sp. contains a mega-array of ANTP-class homeobox genes. Proc Biol Sci. 2023;290:20231563. https://doi.org/10.1098/rspb.2023.1563.
88. Rutlekowski A, Modepalli V, Ketchum R, Moran Y, Reitzel A. De-novo Genome of the Edwardsiid anthozoan Edwardsia elegans. bioRxiv. 2024;2024–10. https://doi.org/10.1101/2024.10.02.616324
89. Ashwood LM, Elnahriry KA, Stewart ZK, et al. Genomic, functional and structural analyses elucidate evolutionary innovation within the sea anemone 8 toxin family. BMC Biol. 2023;21:121. https://doi.org/10.1186/s12915-023-01617-y.
90. Eric E, Kieras M, Pirro S. The Genome Sequences of 118 Taxonomically Diverse Eukaryotes of the Salish Sea. Biodiversity genomes. 2024:2024. https://doi.org/10.56179/001c.118307
91. Li J, Zhan Z, Li Y, Sun Y, Zhou T, Xu K. Chromosome-level genome assembly of a deep-sea Venus flytrap sea anemone sheds light upon adaptations to an extremely oligotrophic environment. Mol Ecol. 2024;33: e17504.
92. Liu C, Bian C, Gao Q, et al. Whole genome sequencing of a novel sea anemone (Actinostola sp.) from a deep-sea hydrothermal vent. Sci Data. 2024;11:102. https://doi.org/10.1038/s41597-024-02944-7.
93. Baumgarten S, Simakov O, Esherick LY, Liew YJ, Lehnert EM, Michell CT, Voolstra CR. The genome of *Aiptasia*, a sea anemone model for coral symbiosis. Proc Natl Acad Sci USA. 2015;112:11893–8. https://doi.org/10.1073/pnas.1513318112.
94. Wood C, Bishop J, Harley J, Mrowicki R, Lab M.B.A.G.A. of Life, W.S.I.T., Darwin Tree of Life Consortium. The genome sequence of the orange-striped anemone, Diadumene lineata (Verrill, 1869). Wellcome Open Research 2022;7. https://doi.org/10.12688/wellcomeopenres.17763.1
95. Feng C, Liu R, Xu W, Zhou Y, Zhu C, Liu J, Wang K. The genome of a new anemone species (Actiniaria: Hormathiidae) provides insights into deep-sea adaptation. Deep-Sea Res I: Oceanogr Res Pap. 2021;170:103492. https://doi.org/10.1016/j.dsr.2021.103492.
96. Zhou Y, Liu H, Feng C, Lu Z, Liu J, Huang Y, Zhang H. Genetic adaptations of sea anemone to hydrothermal environment. Sci Adv. 2023;9:eadh0474. https://doi.org/10.1126/sciadv.adh0474.

97. Adkins P, Bishop J, Mrowicki R, Blaxter ML, Modepalli V, Darwin Tree of Life Consortium. The genome sequence of the brown sea anemone, Metridium senile (Linnaeus, 1761). Wellcome Open Res. 2023;8:536.
98. Aiptasia JBrowse. Reference sequence. http://aiptasia.reefgenomics.org/jbrowse. Accessed 20 October 2024.
99. McFadden CS, Grosberg RK, Cameron BB, Karlton DP, Secord D. Genetic relationships within and between clonal and solitary forms of the sea anemone *Anthopleura elegantissima* revisited: evidence for the existence of two species. Mar Biol. 1997;128:127–39. https://doi.org/10.1007/s002270050076.
100. Glon H, Daly M, Carlton JT, Flenniken MM, Currimjee Z. Mediators of invasions in the sea: life history strategies and dispersal vectors facilitating global sea anemone introductions. Biol Invasions. 2020;22:3195–222. https://doi.org/10.1007/s10530-020-02321-6.
101. LEDAlabShortReadDecontamination. https://github.com/jefferalexdurfue/LEDAlabShortReadDecontamination. Accessed 20 May 2024.
102. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.
103. CAT. https://github.com/MGXlab/CAT_pack. Accessed 20 May 2024.
104. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 20 May 2024.
105. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8. https://doi.org/10.1093/bioinformatics/btw354.
106. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS ONE. 2016;11: e0163962. https://doi.org/10.1371/journal.pone.0163962.
107. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci. 2011;108:1513–8. https://doi.org/10.1073/pnas.1017351108.
108. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, Salzberg SL, Steinegger M. Metagenome analysis using the Kraken software suite. Nat Protoc. 2022;17:2815–39. https://doi.org/10.1038/s41596-022-00738-y.
109. Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. Genome biol. 2020;21:1–31. https://doi.org/10.1186/s13059-020-02154-5.
110. FastqSifter. https://github.com/josephryan/FastqSifter. Accessed 20 May 2024
111. Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 2019;47:10543–52. https://doi.org/10.1093/nar/gkz833.
112. Geneious Prime. https://www.geneious.com. Accessed 02 November 2024.
113. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. Bioinformatics for DNA sequence analysis. 2009;39–64. https://doi.org/10.1007/978-1-59745-251-9_3
114. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 2011;27:171–80. https://doi.org/10.1111/j.1096-0031.2010.00329.x.
115. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9. https://doi.org/10.1038/nmeth.4285.
116. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015.
117. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22. https://doi.org/10.1093/molbev/msx281.
118. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol. 2006;55:539–52. https://doi.org/10.1080/10635150600755453.
119. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst Biol. 2011;60:685–99. https://doi.org/10.1093/sysbio/syr041.
120. Stöver BC, Müller KF. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinform. 2010;11:7. https://doi.org/10.1186/1471-2105-11-7.
121. RyanLabShortReadAssembly. https://github.com/josephryan/RyanLabShortReadAssembly. Accessed 20 May 2024
122. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70. https://doi.org/10.1093/bioinformatics/btr011.
123. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11:1432. https://doi.org/10.1038/s41467-020-14998-3.
124. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T. Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 2014;24:1384–95. https://doi.org/10.1101/gr.170720.113.
125. Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. 2016;44:e113–e113. https://doi.org/10.1093/nar/gkw294.
126. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38:4647–54. https://doi.org/10.1093/molbev/msab199.
127. gVolante. https://gvolante.riken.jp/analysis.html. Accessed 10 October 2024
128. Matemaker. https://github.com/josephryan/matemaker. Accessed 05 May 2024
129. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27:578–9. https://doi.org/10.1093/bioinformatics/btq683.
130. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome biol. 2022;23:258. https://doi.org/10.1186/s13059-022-02823-7.
131. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020;117:9451–7. https://doi.org/10.1073/pnas.1921046117.
132. CD-HIT. https://github.com/weizhongli/cdhit. Accessed 02 July 2024.
133. Seqtk. https://github.com/lh3/seqtk. Accessed 02 July 2024
134. TEsorter. https://github.com/zhangrengang/TEsorter. Accessed 02 July 2024
135. DeepTE. https://github.com/LiLabAtVT/DeepTE. Accessed 02 July 2024
136. DANTE. https://github.com/kavonrtep/dante. Accessed 02 July 2024
137. TEclass2. https://github.com/IOB-Muenster/TEclass2. Accessed 02 July 2024
138. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinform. 2009;25:4–10. https://doi.org/10.1002/0471250953.bi0410s25.
139. dnaPT_utils. https://github.com/clemgoub/dnaPT_utils. Accessed 02 July 2024
140. Durán-Fuentes JA, Maronna MM, Palacios-Gimenez OM, Castillo ER, Ryan JF, Daly M, Stampar SN. Actiniaria-REPlib outputs. Figshare 2025. https://doi.org/10.6084/m9.figshare.27011698.
141. Goubert C, Modolo L, Vieira C, ValienteMoro C, Mavingui P, Boulesteix M. *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*). GBE. 2015;7:1192–205. https://doi.org/10.1093/gbe/evv050.
142. OSC. Ohio Supercomputer Center at The Ohio State University. https://www.osc.edu/. Accessed 10 May 2024.

Publisher's Note