Kon *et al. Mobile DNA* (2025) 16:14
https://doi.org/10.1186/s13100-025-00355-y

Mobile DNA

BRIEF REPORT Open Access

Subtelomeric repeat expansion in *Hydractinia symbiolongicarpus* chromosomes

Tetsuo Kon[1*†], Koto Kon-Nanjo[1†] and Oleg Simakov[1]

Abstract
Despite the striking conservation of animal chromosomes, their repetitive element complements are vastly diverse. Only recently, high quality chromosome-level genome assemblies enabled identification of repeat compositions along a broad range of animal chromosomes. Here, utilizing the chromosome-level genome assembly of *Hydractinia symbiolongicarpus*, a colonial hydrozoan cnidarian, we describe an accumulation of a single 372 bp repeat unit in the subtelomeric regions. Based on the sequence divergence, its partial affinity with the *Helitron* group can be detected. This sequence is associated with a repeated minisatellite unit of about 150 bp. Together, they account for 26.1% of the genome (126 Mb of the 483 Mb). This could explain the genome size increase observed in *H. symbiolongicarpus* compared with other cnidarians, yet distinguishes this expansion from other large cnidarian genomes, such as *Hydra vulgaris*, where such localized propagation is absent. Additionally, we identify a derivative of an IS3EU-like DNA element accumulated at the putative centromeric regions. Our analysis further reveals that *Helitrons* generally comprise a large proportion of *H. symbiolongicarpus* (11.8%). We investigated *Helitron* presence and distributions across several cnidarian genomes. We find that in *Nematostella vectensis*, an anthozoan cnidarian, *Helitron*-like sequences were similarly accumulated at the subtelomeric regions. All these findings suggest that *Helitron* derivatives are prone to forming chromosomal extensions in cnidarians through local amplification in subtelomeric regions, driving variable genome expansions within the clade.

Keywords Transposable elements, *Helitron*, IS3EU, Rolling-circle transposon, Subtelomere, Genome expansion, *Hydractinia symbiolongicarpus*, Cnidarian, Tandem repeats

Introduction

Deep chromosomal homologies have been shown for a vast range of animals and beyond [1, 2]. Despite such conservation of chromosomal identity, the non-coding content and the amount of repetitive elements on these chromosomes are strikingly diverse [3–6]. How repetitive elements are distributed on complete animal chromosomes, has only been uncovered recently by telomere-to-telomere sequence and assembly efforts [7].

Hydractinia symbiolongicarpus is a colonial hydrozoan cnidarian species found along the east coast of North America, forming colonies on hermit crab shells, and can be cultured in laboratory conditions [8]. This species has specialized zooids (polyps) for feeding, reproduction, and defense, and its daily spawning facilitates genetic manipulation techniques like RNAi, CRISPR/Cas9, and transgenesis [9, 10]. *H. symbiolongicarpus* also possesses unique pluripotent interstitial stem cells (i-cells) capable of differentiating into all somatic and germ cell lineages, in contrast to the other commonly studied hydrozoan *Hydra vulgaris* i-cells, which have more limited potential

†Tetsuo Kon and Koto Kon-Nanjo contributed equally to this work.

*Correspondence:
Tetsuo Kon
kontetsuo@gmail.com
[1]Department of Neurosciences and Developmental Biology, University of Vienna, Vienna 1030, Austria

[11]. These features make *H. symbiolongicarpus* an attractive model for studying stem cell function and evolution [9].

In a previous study, we assembled all 15 chromosome sequences of *H. symbiolongicarpus* by combining long-read sequencing with DNase Hi-C [12]. The genome of *H. symbiolongicarpus* retained chromosomal scale synteny with other Cnidaria, although colinear synteny was almost lost compared to *H. vulgaris*, which belongs to the same Hydrozoa class, because of intrachromosomal rearrangements [12, 13]. The genome assembly size was 483 Mb, more than 100 Mb larger than the genome sizes of many other Cnidaria, which typically range from 200 to 300 Mb [14–16]. Although 61% of the genome consisted of repetitive elements, most of them could not be attributed to specific TE families [12]. However, the distribution of sequence divergence in the repetitive elements suggested that there was a past burst of TE activity during a narrow time window [12].

The nucleotide sequences of TEs generally undergo rapid substitutions, and inactive TEs progressively become fragmented, making it often challenging to accurately assign them to known TEs [17–20]. This tendency is particularly evident in animals other than well-studied model organisms such as mice and zebrafish, which have well-characterized genomes [13]. In this study, we used highly sensitive TE analyses, based on the hidden Markov model (HMM) that we previously developed [13]. Our findings suggest that the past TE burst in *H. symbiolongicarpus* was primarily driven by *Helitrons*. Furthermore, by utilizing genome coordinate information obtained from our chromosomal-scale genome assembly [12], we discovered that a single family of repeats potentially derived from *Helitron*-like elements is preferentially accumulated in subtelomeric regions.

Results

TE coverage analysis in the *H. symbiolongicarpus* genome

The previous study suggested that identifying specific TE families responsible for past bursts in *H. symbiolongicarpus* was challenging due to nucleotide substitutions and TE fragmentation [12]. Therefore, in this study, we created a custom repetitive element library using RepeatModeler v2.0.5 [21] with a modified method [13]. This modified method utilized an HMM-based sequence search [22], instead of using rmblast v2.14.1, which is employed by RepeatClassifier v2.0.5, a subprogram of RepeatModeler, to compare the sequences of the custom repetitive element library generated by RepeatModeler against the Dfam database (Fig. S1, Methods) [13, 21]. In the original default run of RepeatModeler v2.0.5, 2,300 sequences were generated as a custom repetitive element library, of which 24.4% (562 sequences) were attributed to one of the TE families. The modified method provided improved annotation: we were eventually able to annotate 63.5% (1,461 sequences) of repeats. Using this refined custom repetitive element library as an input for RepeatMasker v4.1.6, we examined the TE landscape in the chromosome-level genome assembly of *H. symbiolongicarpus* (total size 483 Mb) [12]. As a result, we found that 38.0% (184 Mb) of the genome is occupied by SINEs, LINEs, LTRs, DNA, and rolling-circle (RC) elements (Table S1). When unclassified repetitive elements were also included, the genome coverage expanded to 60.4% (292 Mb) (Table S1). Furthermore, analysis of the nucleotide substitution distribution of TEs (Kimura divergence scores) revealed a sharp peak, as we previously reported [12], suggesting a past burst of TE activities (Fig. 1a, b). Although the main components of the sharp peak were unclear when using the default RepeatModeler custom repetitive element library (Fig. 1a), the analysis with the refined custom repetitive element library showed that *Helitrons* (RC) were the main component in the sharp peak (Fig. 1b). This suggests that the *H. symbiolongicarpus* genome experienced a *Helitron* burst in the past.

We examined the genome coverage of TEs at both the superfamily and family levels [23]. At the superfamily level, the most abundant elements were unclassified elements (22.4%, 108 Mb), *Helitrons* (11.8%, 57.1 Mb), and DNA element IS3EU, (4.19%, 20.2 Mb) (Fig. 1c, Table S2). At the family level, a single family of unclassified element (ID: rnd-1_family-0) occupied 15.9% (76.9 Mb) of the genome (Fig. 1d, Table S3), with an overwhelming copy number of 382,119 copies, far surpassing all other TE families (Fig. 1e, Table S3). We named this unclassified element HSymSR (*H. symbiolongicarpus* subtelomeric repeat) because of its localization to the subtelomeric regions. The HSymSR represented 71.0% of all unclassified elements (Tables S2, S3). A single *Helitron*-like family (ID: rnd-2_family-1) accounted for 10.2% (49.1 Mb) of the genome (Fig. 1d, S2a), representing 86.1% of all *Helitron* elements (Tables S2, S3). We named this *Helitron*-like family HSymSHlR (*H. symbiolongicarpus* subtelomeric Helitron-like repeat). The copy number of the HSymSHlR was 4,938 (Table S3). A single IS3EU-like family (rnd-3_family-1290) accounted for 3.96% (19.1 Mb) of the genome (Figs. 1d, S2b), representing 94.4% of all IS3EU elements (Tables S2S3). We named this IS3EU-like family HSymIS3EUlR (*H. symbiolongicarpus* IS3EU-like repeat). The copy number of the HSymIS3EUlR was 1,384 (Table S3). Moreover, the Kimura substitution level of HSymSHlR was 8.86%, closely aligning with the RC peak (Fig. 1b, Table S3). The Kimura substitution level for the HSymSR was 5.85% (Fig. 1b, Table S3). The HSymIS3EUlR showed a Kimura substitution level of 20.6%, which corresponds to the second wide peak (Fig. 1b, Table S3). In summary, we identified HSymSHlR, HSymSR, and HSymIS3EUlR

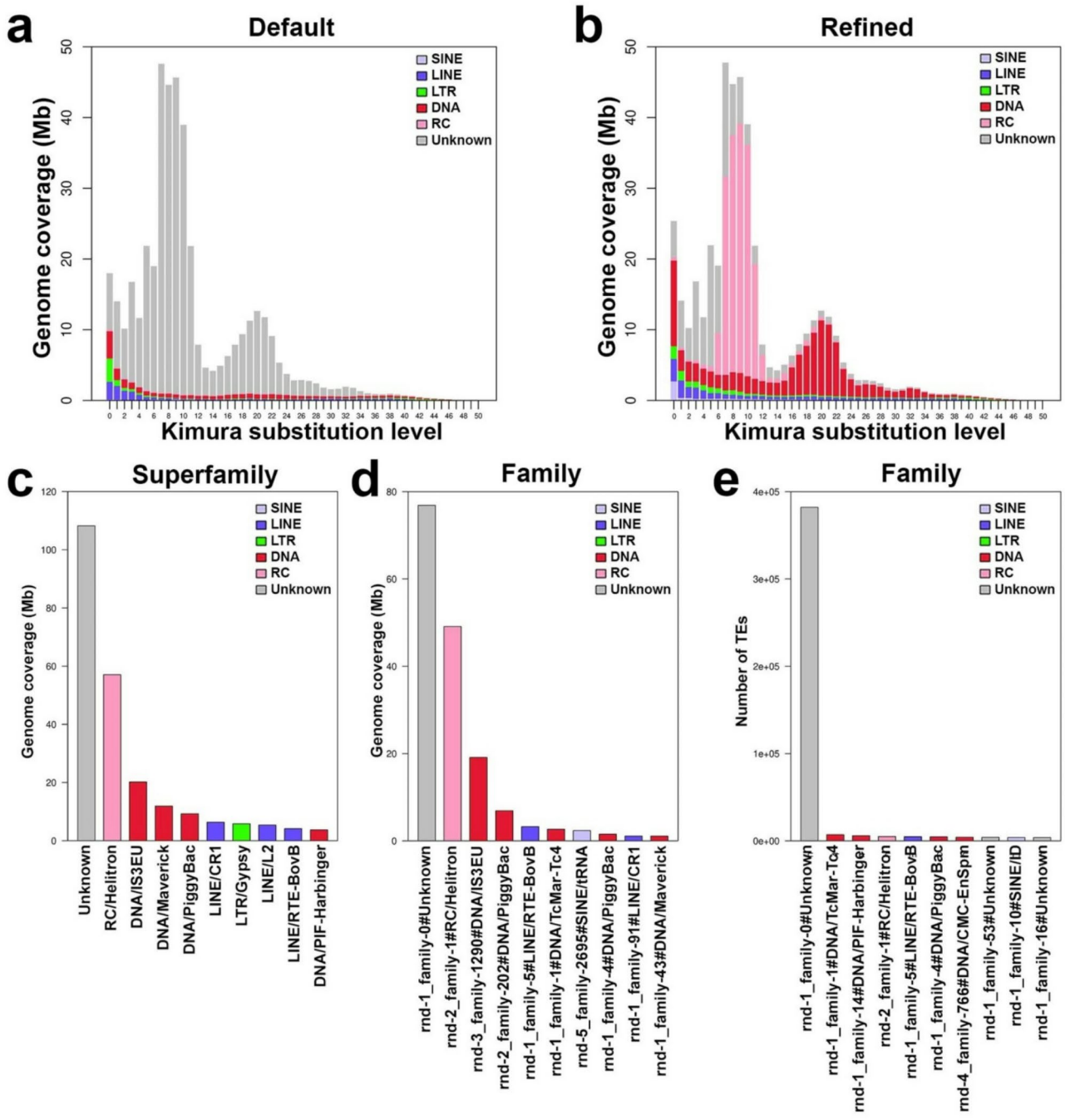

Fig. 1 TE genome coverage in *H. symbiolongicarpus* genome. **(a)** Repeat landscape illustrating TE accumulation history for the *H. symbiolongicarpus* genome with the custom repetitive element library generated by RepeatModeler run with the default settings. The horizontal axis illustrates the Kimura substitution levels of repeat elements relative to their consensus sequences, indicating their ages. On the vertical axis, the graph depicts the genome coverage of each repeat family within the genome in Mb. Consequently, older repetitive elements appear toward the right side of the graph, while more recently active repetitive elements are positioned on the left. (**b**) Repeat landscape with the refined custom repetitive element library. (**c**) Genome coverage of TE superfamilies. (**d**) Genome coverage of TE families. (**e**) Copy number of TE families

that together account for 30.0% (145 Mb) of the genome. These elements may have contributed, at least in part, to the genome size increase in the *H. symbiolongicarpus* genome compared with other cnidarians [14–16].

We also searched for a putative active *Helitron* in our chromosomal-scale genome assembly using the custom repetitive element library. As a result, we identified a 3,389 bp sequence containing two open reading frames (ORFs), which we designated as HSymHel (*H.*

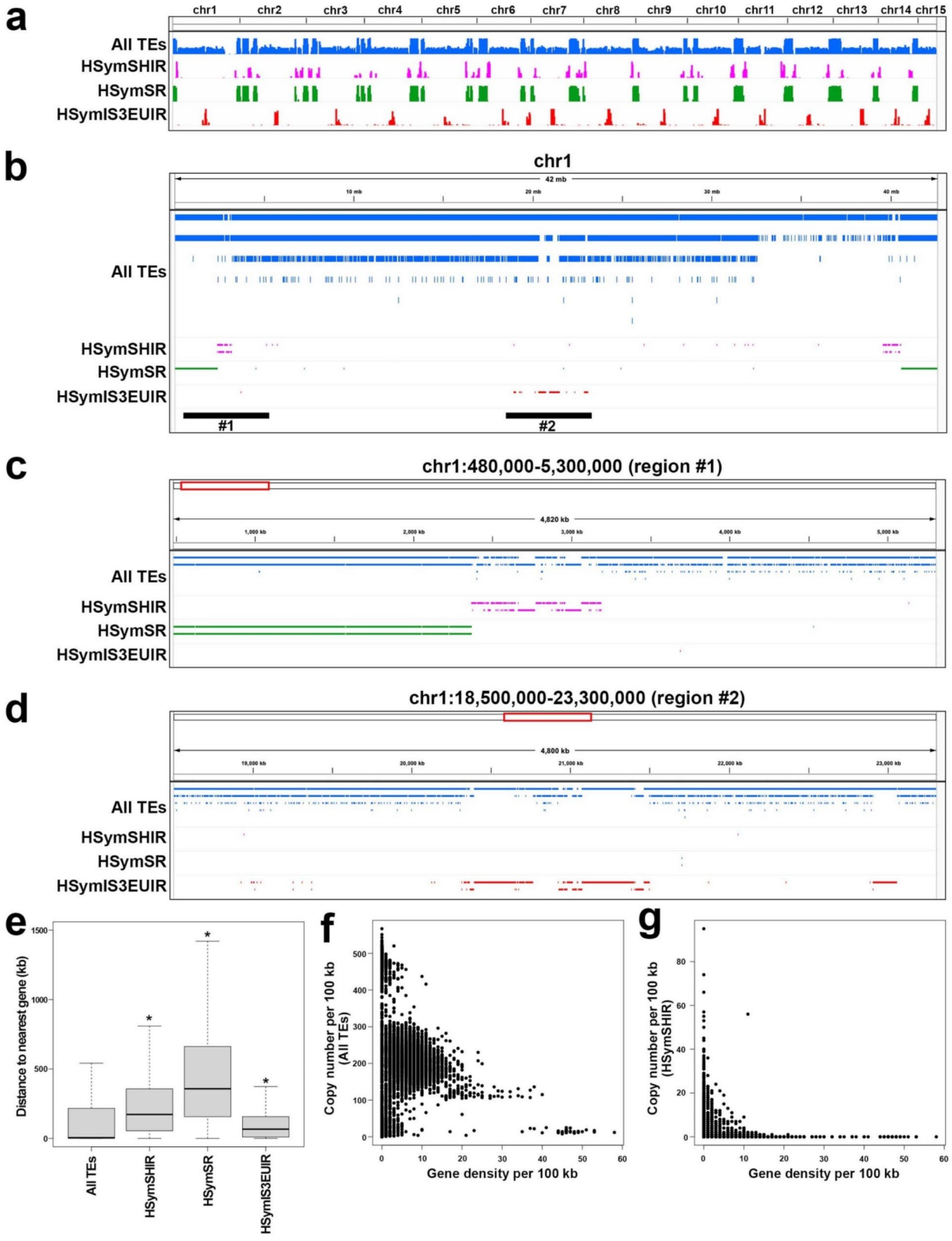

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Subtelomeric and centromic expansions on *H. symbiolongicarpus* chromosomes. (**a**) Whole genome landscape of selected repeat localization. The first track represents all TEs, the second track represents HSymSHIR, and the third track depicts HSymSR. The fourth track shows HSymIS3EUIR. (**b**) Selected repeat localization on the chromosome 1. The regions highlighted by black horizontal lines (#1 and #2) are further emphasized in panel (c) and (d). (**c**) The regional plot on the chromosome 1 (region #1). (**d**) The regional plot on the chromosome 1 (region #2). The first to fourth tracks in panel (b), (c), and (d) are the same as in panels (a). **(e)** Distance to nearest gene. The boxplots illustrate the distance from the nearest gene for all TEs, HSymSHIR, HSymSR, and HSymIS3EUIR. The asterisks indicate that the median distance from the nearest gene is significantly greater compared to all TEs (p-value < 2.2e-16 for all asterisks, Wilcoxon rank-sum test). **(f)** The relationship between gene density and density of all TEs. **(g)** The relationship between gene density and HSymSHIR density

symbiolongicarpus Helitron; Fig. S3a). Analysis of the protein domains encoded by HSymHel revealed that it contains the Helitron_like_N superfamily (NCBI CDD accession cl16715), the DEAD-like_helicase_N superfamily (NCBI CDD accession cl28899), and the RecD superfamily (NCBI CDD accession cl33920) (Fig. S3b). A BLAST search of the HSymHel sequence against the genome mapped it to 217 locations. Among these, 27 sequences that formed alignments longer than 2.5 kb were extracted along with 2 kb upstream and downstream flanking regions, and a multiple alignment was constructed from these 27 sequences and the consensus sequence of HSymHel (Fig. S3c). This analysis demonstrated the conservation of regions corresponding to the three identified domains (Fig. S3c). These findings suggest that HSymHel may represent a putative active *Helitron* in the genome. *Helitrons* can be classified into *Helitron*-like element 1 (HEL1) and *Helitron*-like element 2 (HEL2): HEL1 inserts between 'A' and 'T' in the genome, while HEL2 inserts between 'T' and 'T' [24–28]. By examining the termini of the alignment, we identified potential regions suggesting that HSymHel belongs to HEL1, despite extensive nucleotide substitutions (Fig. S3d). The self-sequence alignment of the HSymHel consensus sequence revealed that it does not contain internal repetitive sequences (Fig. S3e).

Distribution of HSymSHIR copies on chromosomes

Next, we examined how HSymSHIR is distributed on chromosomes. We found that, in all chromosomes except chromosome 15, the HSymSHIR sequences are accumulated at both ends of the chromosomes (Fig. 2a–c, S4, S5). In the regions of the local accumulation, HSymSHIR showed tandem array with variety of unit lengths (Fig. 2b, c). We generated a self-sequence alignment of the consensus sequence of HSymSHIR and found the repetitive nature of the consensus sequence (Fig. S5b). The length of the internal repeat unit of the HSymSHIR consensus sequence is 372 bp (Fig. S5b). Copy numbers and genome coverage of HSymSHIR vary between chromosome arms and median copy number was 148 copies and median genome coverage was 1.46 Mb (Fig. S5c, d). We also observed a tandem array of HSymSR, beginning from the terminal of the array of HSymSHIR and extending continuously to the end of the chromosome (Fig. 2a–c, S5a). Self-sequence alignment of the HSymSR consensus sequence also revealed its repetitive nature (Fig. S5e). The copy numbers and genome coverage of the HSymSR differed across chromosome arms, with a median of 12,278 copies and a median genome coverage of 2.46 Mb (Fig. S5f, g). There was a statistically significant positive correlation between the copy number of HSymSHIR in each chromosome arm and the copy number of HSymSR in each chromosome arm ($p < 9.6 \times 10^{-4}$; Fig. S5h), as well as between the genome coverage of HSymSHIR in each chromosome arm and the genome coverage of HSymSR in each chromosome arm ($p < 1.2 \times 10^{-5}$; Fig. S5i). No statistically significant Blast hits were obtained between the consensus sequence of HSymSHIR and that of HSymSR. We also found that HSymIS3EUIR was highly accumulated at the center of each chromosome and at the right end of chromosome 6 (Fig. 2a, d, S4). In addition to the IS3EU putative homologous region of HSymIS3EUIR (Fig. S2b), the consensus sequence had additional internal repetitiveness with a repeat unit length of 182 bp as determined by Tandem Repeats Finder v4.09. Manual inspection of the raw HiFi genomic reads was able to confirm this consensus sequence.

We also compared the distance to the nearest gene for HSymSHIR, HSymSR, and HSymIS3EUIR against all TEs (Fig. 2e). In all cases, the distance to the nearest gene was significantly greater for HSymSHIR, HSymSR, and HSymIS3EUIR compared to all TEs (Wilcoxon test; $p < 2.2 \times 10^{-16}$ for all cases; Fig. 2e). Furthermore, we analyzed the relationship between gene density and TE copy number in 100 kb genomic windows. Compared to all TEs (Fig. 2f), the TE-rich regions of HSymSHIR (Fig. 2g), HSymSR (Fig. S4b), and HSymIS3EUIR (Fig. S4c) were mutually exclusive with gene-dense regions. These data are consistent with the accumulation of HSymSHIR and HSymSR in subtelomeric regions and HSymIS3EUIR in putative centromeric regions.

For cross-validation, we analyzed another genome assembly of *H. symbiolongicarpus* (Hsym_primary_v1.0) [29]. The contig-level genome assembly of *H. symbiolongicarpus* [29] was organized into 15 pseudochromosomal sequences using the publicly available female genetic linkage map [30]. Then, the TE landscape was examined in these 15 pseudochromosomes. The subtelomeric accumulation of HSymSHIR was also detected in the pseudochromosomes (Fig. S6a, b). The subtelomeric accumulation of HSymSR and the centromeric

accumulation of HSymIS3EUlR were not as distinct as in the chromosome-level genome assembly (Figs. 2, S4, S6a, b). This is most likely due to the nature of pseudochromosomes, where contigs were simply appended to female linkage groups. This could lead to the relatively low resolution in the subtelomeric and centromeric regions.

Molecular phylogenetic analysis of *Helitrons* in the genome of *H. symbiolongicarpus*

Molecular phylogenetic analysis was conducted for copies of HSymSHlR and HSymSR detected in the *H. symbiolongicarpus* genome in order to resolve their phylogenetic relationship. Examining the sequence length distribution of the 4,938 copies from HSymSHlR revealed a broad distribution, with a median length of 2,149 bp (Fig. 3a). In contrast, the 382,119 copies of HSymSR exhibited a narrow length distribution, with a median length of 209 bp (Fig. 3b). Total 177 HSymSHlR sequences with lengths between 2.3 kb and 2.6 kb were extracted and aligned, resulting in a 3,419 bp alignment (Figs. 3c, S7a). Numerous insertions, deletions, and base substitutions were observed, suggesting that these sequences represent inactivated *Helitron* derivatives (Fig. S7a). Similarly, 200 sequences of HSymSR with lengths between 205 bp and 215 bp were randomly extracted and aligned, yielding comparable result (Figs. 3d, S7b). Furthermore, a molecular phylogenetic analysis using the neighbor-joining method was performed on the extracted sequences of HSymSHlR (Fig. 3e). This analysis identified monophyletic groups consisting of copies located on the same chromosome arm, suggesting their local propagation (Fig. 3e).

Genomic distribution of *Helitrons* in other cnidaria

We investigated the presence of *Helitron* sequences in the genome assemblies of five additional cnidarian species (*Hydra vulgaris* [13], *Hydra viridissima* [14], *Rhopilema esculentum* [15], *Acropora millepora* [31], and *Nematostella vectensis* [16]). *Helitrons* were identified in all species analyzed (Fig. 4a). Notably, the genome coverage

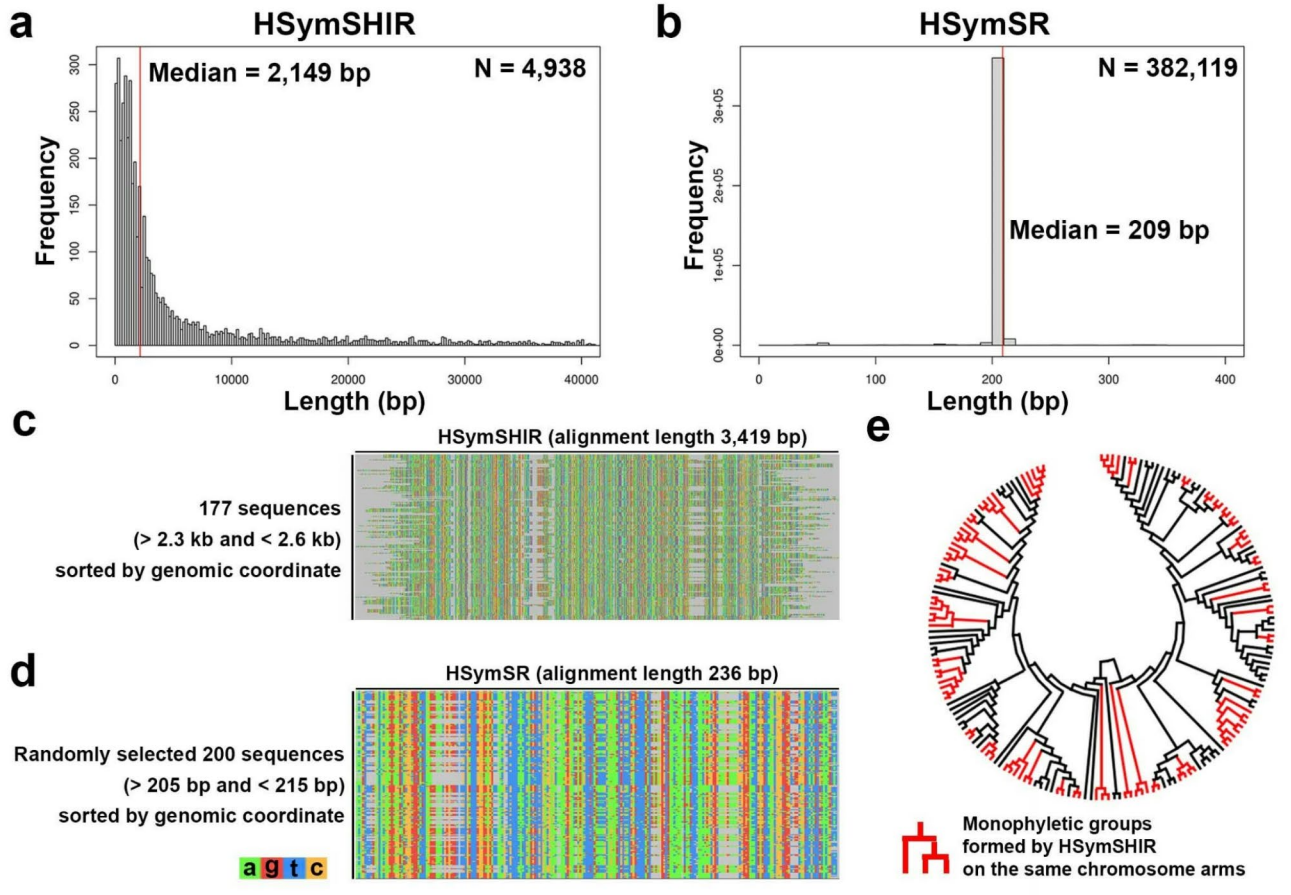

Fig. 3 Molecular phylogenetic analysis of HSymSHlR in *H. symbiolongicarpus* genome. (**a**) Sequence length distribution of HSymSHlR. (**b**) Sequence length distribution of HSymSR. (**c**) Multiple sequence alignment of all 177 HSymSHlR, with sequence lengths greater than 2.3 kb and less than 2.6 kb. (**d**) Multiple sequence alignment of 200 randomly selected HSymSR, with sequence lengths greater than 205 bp and less than 215 bp. The enlarged views of panels (c) and (d) are shown in Fig. S7. (**e**) Cladogram of a neighbor-joining tree of sequences from panel (c). Monophyletic groups formed by HSymSHlR sequences from the same chromosome arms are highlighted in red

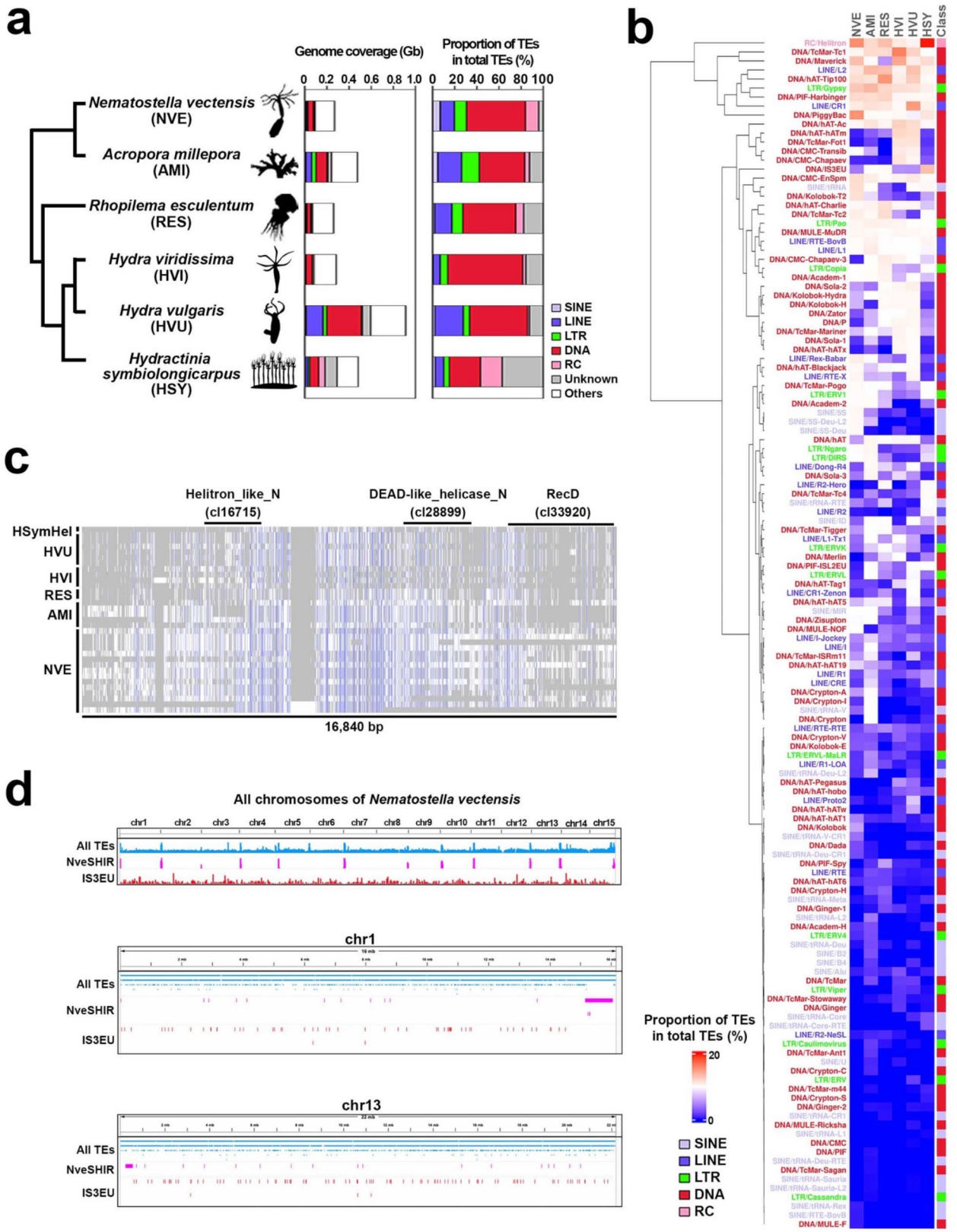

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 *Helitrons* in cnidarian genomes. (**a**) Cladogram of the six cnidarian species with their TE genome coverage and proportion of TEs in total TEs. (**b**) Proportion of TEs in total TEs across the six cnidarian species, with rows representing TEs, sorted by hierarchical clustering using the Euclidean distance and the Ward D2 methods. Columns represent species, all of which show varying amounts of *Helitrons*. (**c**) Sequence alignment of sequences similar to HSymHel from six cnidarians. The domain names are based on the analysis output from the NCBI Conserved Domain Search. (**d**) Genomic localization of *Helitrons* in *Nematostella vectensis*, highlighting the abundance of NveSHlR in subtelomeric regions. The first track shows the distribution of all TEs, the second track shows the distribution of NveSHlR, and the third track shows the distribution of IS3EU. Abbreviations of species names are as follows: NVE, *Nematostella vectensis*; AMI, *Acropora millepora*; RES, *Rhopilema esculentum*; HVI, *Hydra viridissima*; HVU, *Hydra vulgaris*; HSY, *Hydractinia symbiolongicarpus*

of *Helitrons* was higher in *N. vectensis* and *H. symbiolongicarpus* compared to the other species (Fig. 4a, b). Furthermore, we identified sequences similar to HSymHel in these genome assemblies (Fig. 4c). In the generated alignment, although nucleotide substitutions and indels were extensive among species, we found that the three protein domains encoded by HSymHel, such as the Helicase domain, were also present in sequences similar to HSymHel from the five species (Fig. 4c).

Among the analyzed cnidarians, *N. vectensis* exhibited an accumulation of a *Helitron*-like sequence (NveSHlR) near the chromosomal termini, a pattern also observed in *H. symbiolongicarpus* (Fig. 4d, upper panel). These regions were characterized by an extremely low gene density compared to other genomic regions (Fig. 4d, middle and bottom panels). In contrast, IS3EU-like elements, which were found to accumulate in the putative centromeric regions of *H. symbiolongicarpus* (Figs. 2, S2), displayed a diffuse distribution across the chromosomes in *N. vectensis* (Fig. 4d).

We also examined whether canonical (TTAGGG)n telomeric repeats are present at the chromosomal termini of *H. symbiolongicarpus* and *N. vectensis*. As a result, in *H. symbiolongicarpus*, canonical telomeric repeats were identified at one end of chromosome 7. In *N. vectensis*, canonical telomeric repeats were detected at one end of chromosomes 1, 2, 4, 6, 7, and 8. We also performed a BLASTP search using the protein sequence of human telomerase reverse transcriptase (TERT) as a query against the protein sequence sets of both species. As a result, we identified sequences with similarity to TERT in both species (XP_057313859.1, XP_032223148.2), suggesting putative TERT candidates (Fig. S8).

Discussion

Helitrons have been identified across diverse animal and plant species [32–37], with strong evolutionary association with telomeres [38] or centromeres [39]. In the genomes of several plants, such as *Arabidopsis thaliana*, it has been reported that *Helitrons* form tandem arrays within the genome, which is probably a result of rolling-circle replication generating tandem *Helitron* concatemers [40]. Additionally, in *Brassicaceae* species, *Helitrons* are found in lower densities in areas of the genome with high gene density, and they tend to preferentially insert into regions near centromeres and intergenic regions [39]. These findings have primarily been derived from analyses of plant genomes, leaving much to be explored regarding the role of RC transposons in metazoan genomes. In animal genomes, particularly in mammals, *Helitrons* seem to be less common and less active in comparison to plant genomes [41]. As an exceptional example of *Helitron* presence in mammals, in the bat *Myotis lucifugus Helitrons* make up approximately 6% of the genome, and in several cases, they include one or two gene fragments [35]. Likewise, in *Drosophila yakuba*, 3% of the genome consists of both autonomous and non-autonomous Helitrons, including DNAREP1_Dyak, which has been actively transposing within the last 5 million years [42]. Due to the sequence divergence of *Helitrons*, identifying them within repetitive element sequences requires highly sensitive detection methods [13]. Indeed, in our current study, running RepeatModeler with default parameters proved insufficient for detecting *Helitrons* in the *H. symbiolongicarpus* genome. With the modified RepeatModeler pipeline, we observed subtelomeric accumulations of *Helitron*-like elements in *H. symbiolongicarpus* and *N. vectensis*. By applying this approach to other animal genomes, we may be able to shed light on the potentially underestimated role that *Helitrons* could play in metazoan genomes, offering new insights into their evolutionary significance.

The accumulation of Helitron-like elements at the chromosome ends of *H. symbiolongicarpus* raises the possibility that they may contribute to chromosome maintenance. In *Drosophila* species, telomerase has been lost, and to counteract the inevitable telomere shortening, *Drosophila* developed a compensatory mechanism involving certain TEs, primarily non-LTR retrotransposons such as *HeTA*, *TART*, and *TAHRE* [43]. These TEs preferentially insert into telomeric regions, effectively maintaining telomere length in the absence of telomerase [43]. Interestingly, in *Drosophila biarmipes*, Helitrons have been reported to associate with telomeres instead of non-LTR retrotransposons [38]. Although both canonical telomeric repeats and telomerase are present in the genomes of *H. symbiolongicarpus* and *N. vectensis*, the accumulation of repetitive elements at chromosome ends in these species may partially parallel the *Drosophila* system and could potentially contribute to telomere stability. Further studies are needed to evaluate whether these repetitive elements play a functional role in chromosome end protection.

In this study, among the six analyzed cnidarian species (*H. symbiolongicarpus*, *Hydra vulgaris*, *Hydra viridissima*, *Rhopilema esculentum*, *Acropora millepora*, and *N. vectensis*), accumulations of IS3EU-like DNA elements were found in the putative centromeric regions of the *H. symbiolongicarpus* genome. This superfamily was discovered in 2012, with the first ten families of autonomous IS3EU transposons identified in various species (*Danio rerio*, *H. vulgaris*, beetle, acorn worm, and fungus) [44, 45]. IS3EU is a DNA transposon belonging to the IS3 family [46]. Although multiple sequences of IS3EU are registered in Repbase Reports [47, 48], detailed analyses focusing on their structure and function remain largely unexplored, leaving room for future investigation. Unlike *Helitrons*, no significant expansion of IS3EU was detected in *N. vectensis*, suggesting that the proliferation of IS3EU-like elements is specific to the *H. symbiolongicarpus* lineage.

The availability of a chromosome-scale genome assembly for *H. symbiolongicarpus* played a pivotal role in enabling a comprehensive characterization of the distribution of *Helitrons* within the genome. While it is feasible to analyze the average proportion of TEs across the genome using scaffold- or contig-level assemblies, this study revealed that *Helitron*-like elements significantly enrich at the ends of chromosomes, contributing to an increase in genome size. Recent research on other TEs has shown that Ty3/Gypsy elements are highly overrepresented in pericentromeric and subtelomeric regions in frogs using a high-quality chromosome-scale genome assembly [49]. These findings offer valuable insights into the role of TEs in genome evolution, highlighting the importance of high-quality chromosome-scale genome assemblies for detecting such patterns. Thorough investigations of chromosome-scale genome assemblies across various species could uncover novel TE distribution patterns that are dependent on specific genomic regions. This, in turn, may illuminate the role of TEs in maintaining chromosomal integrity, dynamics, and evolution.

In this study, the accumulation of *Helitron*-like *elements* was observed not only in the *H. symbiolongicarpus* genome but also in the *N. vectensis* genome. *H. symbiolongicarpus* and *N. vectensis* belong to Medusozoa and Anthozoa, respectively, both of which are two major clades of Cnidaria that diverged over 500 million years ago [50, 51]. Building on these findings, our study suggests the evolutionary role of *Helitron*-like elements in cnidarian genomes, particularly in shaping the telomeric structure. The conserved chromosomal-level homology observed across cnidarian species, coupled with the accumulation of *Helitron*-like elements in distinct genomic regions of both Medusozoa and Anthozoa lineages, suggests the potential influence of *Helitron*-like elements in Cnidaria chromosome evolution over vast evolutionary timescales. As high-quality chromosomal assemblies continue to become available for cnidarian species [12, 13, 15, 52], further exploration of TE dynamics across diverse lineages will help to identify the ancestral chromosomal characteristics. Together with other eukaryotic studies on *Helitron* distribution [53] this will shed light on broader patterns of *Helitron*-driven genomic innovation in metazoans.

Methods

Generation of custom repetitive element libraries and TE detection in genomes

Annotation of TEs in genomes can sometimes be challenging, largely depending on the extent to which the transposons of the target species or its close relatives are represented in known databases such as Dfam [54]. This difficulty arises because most TEs become inactivated and degraded over time due to the accumulation of mutations, including base substitutions, insertions, and deletions [19, 20]. In such cases, RepeatModeler generates a large number of "Unknown" sequences. Consequently, using a custom repetitive element library containing many such "Unknown" sequences for genome annotation with RepeatMasker [55] would similarly result in a large number of regions being classified as "Unknown" [13]. To overcome this, we adopted an approach similar to that used in our previous study (Fig. S1) [13]. RepeatModeler internally performs iterative sampling and clustering of genomic sequences to generate consensus sequences of repetitive elements (Step 1 in Fig. S1). These consensus sequences are generated solely based on their sequence repetitiveness. In the final step of RepeatModeler, RepeatClassifier annotates them using a homology-based approach by performing a BLAST search against databases such as Dfam (Step 2-A in Fig. S1) [21]. To enhance the sensitivity of this annotation step, we replaced the BLAST-based RepeatClassifier search (Step 2-A in Fig. S1) with an HMM-based search using nhmmscan v3.4 (Step 2-B in Fig. S1) [22]. Applying RepeatMasker with this refined library is anticipated to improve the annotation of genomic regions with identifiable TE classifications (Step 3-B in Fig. S1), compared to using a repetitive library annotated by RepeatClassifier (Step 3-A in Fig. S1).

To implement the refined method, we first obtained the RefSeq *H. symbiolongicarpus* genome sequence from the NCBI Genome database (NCBI RefSeq accession GCF_029227915.1) [12] and generated a custom repetitive element library using RepeatModeler v2.0.5 [21] with the default parameters and the genome sequence as an input. Next, we downloaded all Dfam HMM profiles of TEs from the Dfam v3.8 database (https://dfam.org/home). Due to the large size of the Dfam HMM profile dataset for TEs, we divided it into smaller chunks, each

containing 25,000 HMM profiles. Each chunk file was indexed using hmmpress v3.4 with default parameters [22]. We then used the sequences of the custom repeat library as queries to perform a HMM-based search against the Dfam HMM profiles using nhmmscan v3.4 with default parameters [22]. For each sequence in the custom repetitive element library, we identified the hit sequences in the Dfam database with the lowest E-value, as long as it was less than 0.05. Thereby we refined the annotations in the custom repeat library based on the search results. Using the refined custom repetitive element library, we ran RepeatMasker v4.1.6 [55] on the genome sequences with the options of "-parallel 70 -gff -a -dir -xsmall". For *Hydra vulgaris* (GCA_029227915.2), *Hydra viridissima* (GCA_014706445.1), *Rhopilema esculentum* [15], *Acropora millepora* (GCF_013753865.1), and *N. vectensis* (GCF_932526225.1), we retrieved genome sequences from the NCBI Genome database and performed repetitive element analyses in the same manner as for *H. symbiolongicarpus*.

To investigate the types of repetitive elements present in the genome assembly and their distribution across different genomic regions, we parsed the ".out" file from RepeatMasker v4.1.6 using awk. The ".out" file contains essential information, including the query sequence name in the fifth column, the start and end positions in the sixth and seventh columns, the ID of the repetitive element in the tenth column, and the type of TEs in the eleventh column. We extracted this information using awk and converted it into BED format. Since this BED file represents the genome-wide distribution of TEs, we used it as a basis for downstream analyses. Gene annotation of each species was retrieved from the NCBI Genome database. The positions of genes and TEs on the genomes were visualized using Integrative Genomics Viewer (IGV) v2.12.3 [56].

For the quantitative analysis of the genomic distribution of HSymSHlR, HSymSR, and HSymIS3EUlR, we first prepared BED files containing the distribution of each element, which were generated by parsing the ".out" file from RepeatMasker v4.1.6 using awk. Additionally, we created a gene body BED file from the *H. symbiolongicarpus* genome annotation [12] using awk. Next, we used bedtools closest v2.30.0 [57] with the "-d" option to identify the nearest gene for each TE and calculate the distance to it. We then performed statistical tests using the wilcox.test function of R v4.4.2 [58], considering a p-value < 0.05 as statistically significant, and visualized the results using the R boxplot function. Furthermore, we applied a 100 kb sliding window, shifting by 50 kb across each chromosome, to calculate both the copy number of TEs and the gene density within each window. The results were plotted using R's plot function. The correlation between the copy numbers of HSymSHlR and HSymSR in each chromosome arm was tested using R cor.test function, considering a p-value < 0.05 as statistically significant. The same approach was applied to genome coverage.

Identification of putative active *Helitrons* in the *H. symbiolongicarpus* genome

To identify putative active *Helitrons* in the *H. symbiolongicarpus* genome, we extracted sequences annotated as RC/Helitron from a custom repetitive library generated using RepeatModeler v2.0.5 [21]. Each extracted sequence was then analyzed using the NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to identify sequences encoding *Helitron*-related helicases. This process led to the identification of the HSymHel consensus sequence. The HSymHel consensus sequence was manually curated following the methodology proposed in a previous study [19]. Next, we used the HSymHel consensus sequence as a query to search for related sequences within the *H. symbiolongicarpus* genome using blastn v2.16.0 + with default parameters. To further analyze these sequences, we extended the blast hit coordinates by 2 kb upstream and downstream and defined these regions using a BED format file. The corresponding genomic sequences were then extracted using bedtools getfasta v2.30.0 [57]. The extracted sequences were aligned using Clustal Omega v1.2.4 [59], and the resulting alignment was visually inspected in Jalview v2.11.4.1 [60]. Additionally, a self-dotplot of HSymHel was generated using the NCBI BLAST server to examine the presence of internal repeat structures within the sequence. The length of the internal repeat unit of the consensus sequences was estimated using Tandem Repeats Finder v4.09 with the options "2 7 7 80 10 50 500" (https://tandem.bu.edu/trf/trf.html).

Analysis using an alternative genome assembly of *H. symbiolongicarpus*

An alternative genome assembly of *H. symbiolongicarpus* (Hsym_primary_v1.0 [29]) was obtained from the Hydractinia Genome Project Portal (https://research.nhgri.nih.gov/hydractinia/). Because the Hsym_primary_v1.0 is a contig-level genome assembly, we generated a pseudochromosomal sequences by anchoring the contig sequences of the Hsym_primary_v1.0 assembly to the previously reported female genetic linkage maps of *H. symbiolongicarpus* [30]. In this process, we aligned the contig sequences based on the SNP marker positions [30] and constructed 15 pseudochromosome sequences. The TE landscape on the 15 pseudochromosome sequences was analyzed using RepeatMasker v4.1.6 with a custom repetitive element library generated from the RefSeq *H. symbiolongicarpus* genome assembly

(NCBI RefSeq accession GCF_029227915.1) following the method described above.

TE distribution analyses

For the analysis of Kimura substitution levels, we extracted the Kimura substitution values for each TE from the RepeatMasker output file and visualized them using the barplot function in R v4.4.2. Similarly, the genome-wide coverage of TEs was calculated based on data extracted from the RepeatMasker result. The proportions of various TE contents were visualized as a heatmap using the ComplexHeatmap R library v2.20.0 with default parameters [61]. For generating the self-alignments of consensus sequences of TE families, alignments were generated using blastn v2.16.0 + with default parameters [62]. The alignments were visualized using the blast2dotplot.py script (commit hash: 3b55d11) from the bio_small_scripts repository (https://github.com/satoshikawato/bio_small_scripts) with default parameters. DNA sequence multiple alignments were generated using Clustal Omega v1.2.4 [59] with its default parameters. The resulting multiple sequence alignments were visualized using Jalview v2.11.4.1 [60]. Molecular phylogenetic analyses were performed using MEGA software v11 [63].

Identification of sequences similar to HSymHel in other cnidarian genomes

To identify sequences similar to HSymHel in the genomes of other Cnidarians, we conducted tblastn v2.16.0 + searches using the protein sequence of ORF1 from HSymHel as the query against RepeatModeler libraries generated from the genome assemblies of *H. vulgaris* [13], *H. viridissima* [14], *R. esculentum* [15], *A. millepora* [31], and *N. vectensis* [16]. The search was performed using default parameters, and we identified top-hit elements that were also annotated as *Helitron*. Next, these identified sequences were mapped to their respective genome assemblies using blastn v2.16.0 + with default parameters. Only regions with alignments longer than 2.5 kb in *N. vectensis* and *A. millepora*, 900 bp in *H. viridissima*, and 1.6 kb in *H. vulgaris* were retained. We then extended these regions by 2 kb upstream and downstream, defining them in a BED file, and extracted the corresponding sequences from the genome using bedtools getfasta v2.30.0 with default parameters. For *R. esculentum*, no direct hits were found in the repeat library, so we performed a direct tblastn v2.16.0 + search against the genome. Regions with alignments longer than 500 bp were selected, and the flanking 2 kb upstream and downstream regions were extracted. Finally, all extracted sequences were aligned using Clustal Omega v1.2.4, and the alignment was visualized and inspected in Jalview v2.11.4.1.

Abbreviations

AMI	*Acropora millepora*
HEL1	*Helitron*-like element 1
HEL2	*Helitron*-like element 2
HMM	Hidden Markov Model
HSY	*Hydractinia symbiolongicarpus*
HSymHel	*H. symbiolongicarpus* Helitron
HSymSHIR	*H. symbiolongicarpus* subtelomeric *Helitron*-like repeat
HSymIS3EUIR	*H. symbiolongicarpus* IS3EU-like repeat
HSymSR	*H. symbiolongicarpus* subtelomeric repeat
HVU	*Hydra vulgaris*
HVI	*Hydra viridissima*
NveSHIR	*N. vectensis* subtelomeric *Helitron*-like repeat
NVE	*Nematostella vectensis*
RC	Rolling-Circle
RCR	Rolling-Circle Replication
RES	*Rhopilema esculentum*
TEs	Transposable Elements
TERT	Telomerase Reverse Transcriptase

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13100-025-00355-y.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

The authors thank Uri Frank at the University of Galway for comments on the manuscript, as well as Nina Znidaric, Angela Caballero Alfonso, and Eduard Renfer at the University of Vienna for their technical assistance. The computation for this study was performed at the Life Science Compute Cluster (LiSC) of the University of Vienna. The authors sincerely thank the editor and the two anonymous reviewers for their valuable feedback, which greatly helped improve this manuscript.

Author contributions

TK, KK, and OS designed the study. TK, KK, and OS analyzed the data. TK, KK, and OS wrote the manuscript. All authors reviewed the manuscript and agreed to the content.

Funding

This work was supported by the Austrian Science Fund FWF (I4353) to OS, ERC-H2020-EURIP grant (grant number 945026), Takeda Science Foundation to TK, the Mochida Memorial Foundation for Medical and Pharmaceutical Research to TK, the International Medical Research Foundation to TK, the Yamada Science Foundation to TK.

Data availability

The consensus sequences for the HSymSHIR, HSymSR, HSymHel and other related data have been deposited at the Figshare repository under https://doi.org/10.6084/m9.figshare.28533614.v3. All analyses were conducted using publicly available datasets, as detailed in the Methods section, which served as the primary data sources for this research.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 21 November 2024 / Accepted: 17 March 2025
Published online: 25 March 2025

References

1. Simakov O, Bredeson J, Berkoff K, Marletaz F, Mitros T, Schultz DT et al. Deeply conserved synteny and the evolution of metazoan chromosomes. Sci Adv [Internet]. 2022;8:eabi5884. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35108053
2. Schultz DT, Haddock SHD, Bredeson JV, Green RE, Simakov O, Rokhsar DS. Ancient gene linkages support ctenophores as sister to other animals. Nature [Internet]. 2023;618:110–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/37198475
3. Schartl M, Woltering JM, Irisarri I, Du K, Kneitz S, Pippel M et al. The genomes of all lungfish inform on genome expansion and tetrapod evolution. Nature [Internet]. 2024;634:96–103. Available from: https://www.nature.com/articles/s41586-024-07830-1
4. Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I et al. Giant lungfish genome elucidates the conquest of land by vertebrates. Nature [Internet]. 2021;590:284–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33461212
5. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A et al. Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization. Science (80-) [Internet]. 2007;317:86–94. Available from: https://www.science.org/doi/https://doi.org/10.1126/science.1139158
6. Kissinger JC, DeBarry J. Genome cartography: charting the apicomplexan genome. Trends Parasitol [Internet]. 2011;27:345–54. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1471492211000572
7. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA et al. Earth BioGenome Project: Sequencing life for the future of life. Proc Natl Acad Sci U S A [Internet]. 2018;115:4325–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29686065
8. Buss LW, Yund PO, A Sibling Species Group of Hydractinia in the North-Eastern United States. J Mar Biol Assoc United Kingdom [Internet]. 1989;69:857–74. Available from: https://www.cambridge.org/core/product/identifier/S0025315400032215/type/journal_article
9. Frank U, Nicotra ML, Schnitzler CE. The colonial cnidarian Hydractinia. Evodevo [Internet]. 2020;11:7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32226598
10. Chrysostomou E, Febrimarsa, DuBuc T, Frank U. Gene Manipulation in Hydractinia. Methods Mol Biol [Internet]. 2022;2450:419–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35359321
11. Varley Á, Horkan HR, McMahon ET, Krasovec G, Frank U. Pluripotent, germ cell competent adult stem cells underlie cnidarian regenerative ability and clonal growth. Curr Biol [Internet]. 2023; Available from: http://www.ncbi.nlm.nih.gov/pubmed/37028430
12. Kon-Nanjo K, Kon T, Horkan HR, Febrimarsa, Steele RE, Cartwright P et al. Chromosome-level genome assembly of Hydractinia symbiolongicarpus. G3 (Bethesda) [Internet]. 2023;13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/37294738
13. Kon-Nanjo K, Kon T, Koubkova Yu TC-T, Rodriguez-Terrones D, Falcon F, Martínez DE et al. The dynamic genomes of hydra and the anciently active repeat complement of animal chromosomes. BioRxiv. 2024.
14. Hamada M, Satoh N, Khalturin K. A Reference Genome from the Symbiotic Hydrozoan, Hydra viridissima. G3 (Bethesda) [Internet]. 2020;10:3883–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32900905
15. Li Y, Gao L, Pan Y, Tian M, Li Y, He C et al. Chromosome-level reference genome of the jellyfish Rhopilema esculentum. Gigascience [Internet]. 2020;9. Available from: https://academic.oup.com/gigascience/article/doi/https://doi.org/10.1093/gigascience/giaa036/5823175
16. Fletcher C, Pereira da Conceicoa L. The genome sequence of the starlet sea anemone, Nematostella vectensis (Stephenson, 1935). Wellcome Open Res [Internet]. 2023;8:79. Available from: https://wellcomeopenresearch.org/articles/8-79/v1
17. Peona V, Martelossi J, Almojil D, Bocharkina J, Brännström I, Brown M et al. Teaching transposon classification as a means to crowd source the curation of repeat annotation - a tardigrade perspective. Mob DNA [Internet]. 2024;15:10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/38711146
18. Orozco-Arias S, Sierra P, Durbin R, González J. MCHelper automatically curates transposable element libraries across eukaryotic species. Genome Res [Internet]. 2024;34:2256–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/39653419
19. Goubert C, Craig RJ, Bilat AF, Peona V, Vogan AA, Protasio AV. A beginner's guide to manual curation of transposable elements. Mob DNA [Internet]. 2022;13:7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35354491
20. Matsushima W, Planet E, Trono D. Ancestral genome reconstruction enhances transposable element annotation by identifying degenerate integrants. Cell genomics [Internet]. 2024;4:100497. Available from: http://www.ncbi.nlm.nih.gov/pubmed/38295789
21. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A [Internet]. 2020;117:9451–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32300014
22. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics [Internet]. 2013;29:2487–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23842809
23. Kapitonov VV, Jurka J. A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet [Internet]. 2008;9:411–2; author reply 414. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18421312
24. Chellapan BV, van Dam P, Rep M, Cornelissen BJC, Fokkens L. Non-canonical Helitrons in Fusarium oxysporum. Mob DNA [Internet]. 2016;7:27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27990178
25. Heringer P, Kuhn GCS. Pif1 Helicases and the Evidence for a Prokaryotic Origin of Helitrons. Mol Biol Evol [Internet]. 2022;39. Available from: http://www.ncbi.nlm.nih.gov/pubmed/34850089
26. Thomas J, Vadnagara K, Pritham EJ. DINE-1, the highest copy number repeats in Drosophila melanogaster are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). Mob DNA [Internet]. 2014;5:18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24959209
27. Bao W, Jurka J. Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. Mob DNA [Internet]. 2013;4:12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23548000
28. Li Z, Gilbert C, Peng H, Pollet N. Discovery of numerous novel Helitron-like elements in eukaryote genomes using HELIANO. Nucleic Acids Res [Internet]. 2024;52:e79. Available from: http://www.ncbi.nlm.nih.gov/pubmed/39119924
29. Schnitzler CE, Chang ES, Waletich J, Quiroga-Artigas G, Wong WY, Nguyen A-D et al. The genome of the colonial hydroid Hydractinia reveals that their stem cells use a toolkit of evolutionarily shared genes with all animals. Genome Res [Internet]. 2024; Available from: http://genome.cshlp.org/lookup/doi/https://doi.org/10.1101/gr.278382.123
30. Chen R, Sanders SM, Ma Z, Paschall J, Chang ES, Riscoe BM et al. XY sex determination in a cnidarian. BMC Biol [Internet]. 2023;21:32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/36782149
31. Fuller ZL, Mocellin VJL, Morris LA, Cantin N, Shepherd J, Sarre L et al. Population genetics of the coral Acropora millepora: Toward genomic prediction of bleaching. Science (80-) [Internet]. 2020;369. Available from: https://www.science.org/doi/https://doi.org/10.1126/science.aba4674
32. Du C, Fefelova N, Caronna J, He L, Dooner HK. The polychromatic Helitron landscape of the maize genome. Proc Natl Acad Sci U S A [Internet]. 2009;106:19916–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19926866
33. Yang L, Bennetzen JL. Distribution, diversity, evolution, and survival of Helitrons in the maize genome. Proc Natl Acad Sci U S A [Internet]. 2009;106:19922–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19926865
34. Choi J-D, Hoshino A, Park K-I, Park I-S, Iida S. Spontaneous mutations caused by a Helitron transposon, Hel-It1, in morning glory, Ipomoea tricolor. Plant J [Internet]. 2007;49:924–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17257169
35. Thomas J, Phillips CD, Baker RJ, Pritham EJ. Rolling-Circle Transposons Catalyze Genomic Innovation in a Mammalian Lineage. Genome Biol Evol [Internet]. 2014;6:2595–610. Available from: https://academic.oup.com/gbe/article-lookup/doi/https://doi.org/10.1093/gbe/evu204
36. Han M-J, Shen Y-H, Xu M-S, Liang H-Y, Zhang H-H, Zhang Z. Identification and Evolution of the Silkworm Helitrons and their Contribution to Transcripts. DNA Res [Internet]. 2013;20:471–84. Available from: https://academic.oup.com/dnaresearch/article-lookup/doi/https://doi.org/10.1093/dnares/dst024
37. Han G, Zhang N, Xu J, Jiang H, Ji C, Zhang Z et al. Characterization of a novel Helitron family in insect genomes: insights into classification, evolution and horizontal transfer. Mob DNA [Internet]. 2019;10:25. Available from: https://mobilednajournal.biomedcentral.com/articles/https://doi.org/10.1186/s13100-019-0165-4
38. Saint-Leandre B, Nguyen SC, Levine MT. Diversification and collapse of a telomere elongation mechanism. Genome Res [Internet]. 2019;29:920–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31138619

39. Hu K, Xu K, Wen J, Yi B, Shen J, Ma C et al. Helitron distribution in Brassicaceae and whole Genome Helitron density as a character for distinguishing plant species. BMC Bioinformatics [Internet]. 2019;20:354. Available from: https://bmcbioinformatics.biomedcentral.com/articles/https://doi.org/10.1186/s12859-019-2945-8
40. Xiong W, Dooner HK, Du C. Rolling-circle amplification of centromeric Helitrons in plant genomes. Plant J [Internet]. 2016;88:1038–45. Available from: https://onlinelibrary.wiley.com/doi/https://doi.org/10.1111/tpj.13314
41. Barro-Trastoy D, Köhler C. Helitrons: genomic parasites that generate developmental novelties. Trends Genet [Internet]. 2024;40:437–48. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168952524000295
42. Kapitonov VV, Jurka J. Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet [Internet]. 2007;23:521–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17850916
43. Levin HL, Moran JV. Dynamic interactions between transposable elements and their hosts. Nat Rev Genet [Internet]. 2011;12:615–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21850042
44. Kapitonov VV, Bao W, Jurka J. IS3EU-1_DR - a family of autonomous DNA transposon from the zebrafish genome. Direct Submiss to Repbase Updat (25- JUN-2012). 2012.
45. Kapitonov VV, Bao W, Jurka J. Simple DNA transposons: scratching the surface. Direct Submiss to Repbase Updat. 2012.
46. Kojima KK. Structural and sequence diversity of eukaryotic transposable elements. Genes Genet Syst [Internet]. 2019;94:233–52. Available from: https://www.jstage.jst.go.jp/article/ggs/94/6/94_18-00024/_article
47. Bao W, Jurka J. IS3EU-1N1_DR. Repbase Rep. 2014;14.
48. Bao W. IS3EU-1_CorFlu. Repbase Rep. 2023;23.
49. Bredeson JV, Mudd AB, Medina-Ruiz S, Mitros T, Smith OK, Miller KE et al. Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs. Nat Commun [Internet]. 2024;15:579. Available from: https://www.nature.com/articles/s41467-023-43012-9
50. Schwentner M, Bosch TCG. Revisiting the age, evolutionary history and species level diversity of the genus Hydra (Cnidaria: Hydrozoa). Mol Phylogenet Evol [Internet]. 2015;91:41–55. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1055790315001566
51. Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeek MA. Estimating metazoan divergence times with a molecular clock. Proc Natl Acad Sci [Internet]. 2004;101:6536–41. Available from: https://doi.org/10.1073/pnas.0401670101
52. Zimmermann B, Montenegro JD, Robb SMC, Fropf WJ, Weilguny L, He S et al. Topological structures and syntenic conservation in sea anemone genomes. Nat Commun [Internet]. 2023;14:8270. Available from: https://www.nature.com/articles/s41467-023-44080-7
53. Jedlička P, Tokan V, Kejnovská I, Hobza R, Kejnovský E. Telomeric retrotransposons show propensity to form G-quadruplexes in various eukaryotic species. Mob DNA [Internet]. 2023;14:3. Available from: https://mobilednajournal.biomedcentral.com/articles/https://doi.org/10.1186/s13100-023-00291-9
54. 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 [Internet]. 2021;12:2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33436076
55. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinforma [Internet]. 2009;Chap. 4:4.10.1–4.10.14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19274634
56. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G et al. Integrative genomics viewer. Nat Biotechnol [Internet]. 2011;29:24–6. Available from: https://www.nature.com/articles/nbt.1754
57. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics [Internet]. 2010;26:841–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20110278
58. R Core Team. R: A Language and environment for statistical computing. R Found Stat Comput. 2024.
59. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol [Internet]. 2011;7. Available from: https://www.embopress.org/doi/https://doi.org/10.1038/msb.2011.75
60. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics [Internet]. 2009;25:1189–91. Available from: https://academic.oup.com/bioinformatics/article/25/9/1189/203460
61. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics [Internet]. 2016;32:2847–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27207943
62. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol [Internet]. 1990;215:403–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2231712
63. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Battistuzzi FU, editor. Mol Biol Evol [Internet]. 2021;38:3022–7. Available from: https://academic.oup.com/mbe/article/38/7/3022/6248099

Publisher's note