



DISSERTATION | DOCTORAL THESIS

Titel | Title

Activity and evolution of transposable elements along anciently conserved animal chromosomes

verfasst von | submitted by

Koto Kon

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 794 685 437

Dissertationsgebiet lt. Studienblatt | Field of
study as it appears on the student record
sheet:

Biologie

Betreut von | Supervisor:

Univ.-Prof. Dr. Oleg Simakov

Table of contents

Acknowledgements	3
Publication list	4
Abstract.....	5
Zusammenfassung	6
Introduction	8
Results	
Paper I: The dynamic genomes of Hydra and the anciently active repeat complement of animal chromosomes	20
Paper II: Chromosome-level genome assembly of Hydractinia symbiolongicarpus	67
Paper III: Subtelomeric repeat expansion in Hydractinia symbiolongicarpus chromosomes	81
Concluding Discussion.....	104

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Oleg Simakov, for his continuous guidance, encouragement, and generous support throughout my PhD studies. Under his supervision, I have learned not only the fundamentals of conducting rigorous scientific research but also the broader perspectives of thinking critically and creatively. From genomics and Hydra experiments to writing scientific papers and presenting research effectively, his mentorship has shaped every aspect of my growth as a researcher. The education I received under his supervision will remain a lifelong asset. His constant encouragement and trust have been the driving forces that made this thesis possible.

I am also sincerely grateful to my collaborators and mentors, Prof. Thomas W. Holstein (University of Heidelberg, Heidelberg, Germany), Prof. Elly Tanaka (Institute of Molecular Biotechnology, Vienna, Austria), Prof. Rob Steele (University of California, Irvine, USA), Prof. Daniel Martínez (Pomona College, Claremont, USA), and Prof. Uri Frank (University of Galway, Galway, Ireland). Their insightful discussions and constructive feedback have greatly refined the direction of the projects. I deeply appreciate the opportunities to interact and exchange ideas with such inspiring scientists.

I am also thankful to Prof. Bert Hobmayer (University of Innsbruck, Innsbruck, Austria), for his insightful comments on this project. Special thanks to Kitti Dora Csalyi and Thomas Sauer at the Max Perutz Labs BioOptics FACS Facility for their valuable advice on the FACS experiments.

This work was supported by the Austrian Science Fund (FWF, grant I4353), whose financial assistance made this research possible.

Finally, I would like to express my heartfelt gratitude to my family. Thank you for believing in my decision to pursue a PhD and for standing by me every single day. Your understanding and encouragement have sustained me through every challenge and made this journey truly meaningful.

Publication list

Main projects

I. **Kon-Nanjo K**, Kon T, Koubkova Yu CT, 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. *Genome Biology*. 26(1):186, 2025. DOI: 10.1186/s13059-025-03653-z.

II. **Kon-Nanjo K**, Kon T, Horkan HR, Febrimarsa, Steele RE, Cartwright P, Frank U, Simakov O. Chromosome-level genome assembly of *Hydractinia symbiolongicarpus*. *G3: Genes, Genomes, Genetics*. 13(8): jkad107, 2023. DOI: 10.1093/g3journal/jkad107.

III. Kon T#, **Kon-Nanjo K#** (#equally contributed), Simakov O. Expansion of a single Helitron subfamily in *Hydractinia symbiolongicarpus* suggests a shared mechanism of cnidarian chromosomal extension. *Mobile DNA*. 16 (14): 2025. DOI: 10.1186/s13100-025-00355-y.

Subprojects

IV. Kon T, **Kon-Nanjo K**, Syaputri Handayani K, Zang L, Fahrurrozi F, Simakov O, Gultom VDN, Shimada Y. Chromosome-level genome assembly of the doctor fish (*Garra rufa*). *Scientific Data*. 9;12(1):765, 2025. DOI: 10.1038/s41597-025-05101-w.

V. Shimada Y, Aydın B, **Kon-Nanjo K**, Syaputri Handayani K, Gultom VDN, Simakov O, Fahrurrozi, Kon T. Potential of *Garra rufa* as a novel high- temperature resistant model fish: a review on current and future approaches. *Zoological Lett*. 11(1):3. DOI: 10.1186/s40851-025-00249-0, 2025.

VI. Yoshida M, Tóth E, **Kon-Nanjo K**, Kon T, Hirota K, Toyoda A, Toh H, Miyazawa H, Terauchi M, Noguchi H, Setiamarga DHE, Simakov O. Giant genome of the vampire squid reveals the derived state of modern octopod karyotypes. *iScience*. 113832, 2025. DOI: 10.1016/j.isci.2025.113832.

VII. Ahuja N, Schultz DT, Destanović D, Church SH, Picciani N, Munro C, **Kon-Nanjo K**, Kon T, Maňko MK, Shi W, Haddock SHD, Simakov O, Dunn CW. A Chromosome-Scale Genome of *Nanomia septata* Reveals Extensive Rearrangement But No Clear Driver of the Unique Colony-Level Organization of Siphonophores. *bioRxiv* 2025. DOI: 10.1101/2025.05.01.651713.

Abstract

Animal genomes exhibit a remarkable conservation of chromosomal organization, yet they vary extensively in size and repetitive element content. Transposable elements (TEs) are major contributors to this variation and play fundamental roles in shaping genome architecture and evolution. While retrotransposons (Class I) have been implicated in many instances of genome expansion, the contribution of DNA transposons (Class II) and their chromosomal dynamics remain less well understood. Furthermore, how TE activity is maintained and regulated in stem-cell genomes remains an unresolved question in genome biology.

In this thesis, I addressed these issues by combining stem-cell-resolved genomics and comparative genomics analyses using cnidarian model systems *Hydra vulgaris* and *Hydractinia symbiolongicarpus*. These basally branching metazoans possess extraordinary regenerative abilities and provide unique opportunities to study the interplay between genome stability, TE dynamics in stem-cell genomes, and stem-cell functions.

In the first study (Kon-Nanjo *et al*, Genome Biology 2025), I generated haplotype-resolved, telomere-to-telomere genome assemblies of two genetically divergent hydra strains and performed genomic and transcriptomic analyses across the organism's three stem-cell lineages. This work revealed that distinct TE families are transcriptionally and genomically active in each lineage through non-random insertions. A deeply conserved core set of mainly DNA transposons was identified as an anciently active component of animal chromosomes, contributing to gradual genomic expansions across metazoans. These elements influence structural variation around loci associated with cell proliferation and chromatin topology, suggesting that evolutionarily persistent set of transposons acts as self-renewing components of chromosome architecture.

In the second study (Kon-Nanjo *et al*, G3: Genes, Genomes, Genetics, 2023), I produced the first chromosome-level genome assembly of the colonial hydrozoan *Hydractinia symbiolongicarpus* using PacBio HiFi long-read sequencing and Hi-C scaffolding. The 483 Mb assembly, organized into 15 chromosomes representing 99.8% of the genome, revealed that 61% consists of repetitive sequences, reflecting at least two distinct phases of repeat expansion. With 25,825 protein-coding genes (93.1% BUSCO completeness) and extensive macrosynteny conservation with *Hydra vulgaris*, this genome provides a comprehensive platform for comparative analyses of hydrozoan stem-cell biology and chromosomal evolution.

In the third study (Kon-Nanjo *et al*, Mobile DNA, 2025), I uncovered a massive subtelomeric accumulation of a single 372 bp repeat unit related to the Helitron DNA transposon family in *Hydractinia symbiolongicarpus*. Together with a tandem minisatellite repeat, this element family accounts for 26% of the total genome and largely explains its size expansion relative to other cnidarians. Comparative analyses showed that Helitron-derived sequences also accumulate near chromosome ends in other cnidarians, indicating that localized amplification of these elements may drive subtelomeric extension and lineage-specific genome expansion within the phylum.

Together, these studies demonstrate that TEs act as both conserved and dynamic components of animal chromosomes. They shape genome diversification while maintaining large-scale chromosomal stability, thereby bridging the long-standing gap between chromosomal conservation and chromosomal evolutionary dynamics.

Zusammenfassung

Tierische Genome zeigen eine bemerkenswerte Erhaltung der chromosomalen Organisation, unterscheiden sich jedoch erheblich in ihrer Größe und im Anteil repetitiver Elemente. Transposable Elemente (TEs) tragen wesentlich zu dieser Variation bei und spielen eine grundlegende Rolle bei der Gestaltung der Genomarchitektur und der Evolution. Während Retrotransposons (Klasse I) in vielen Fällen der Genomexpansion eine Rolle spielen, sind der Beitrag von DNA-Transposons (Klasse II) und ihre chromosomale Dynamik bislang weniger gut verstanden. Darüber hinaus ist bislang ungeklärt, wie die Aktivität von TEs in Stammzellgenomen aufrechterhalten und reguliert wird – eine zentrale Frage der Genombiologie.

In dieser Dissertation habe ich diese Fragen untersucht, indem ich stammzellaufgelöste Genomanalysen mit vergleichender Genomik kombinierte und *Cnidaria*, darunter *Hydra vulgaris* und *Hydractinia symbiolongicarpus*, als Modellsysteme verwendete. Diese basalen Metazoen besitzen außergewöhnliche Regenerationsfähigkeiten und bieten einzigartige Möglichkeiten, das Zusammenspiel von Genomstabilität, TE-Dynamik in Stammzellgenomen und Stammzellfunktionen zu erforschen.

In der ersten Studie (Kon-Nanjo et al., *Genome Biology*, 2025) erstellte ich haplotypaufgelöste, telomer-zu-telomer-Genomassemblierungen von zwei genetisch divergenten *Hydra*-Stämmen und führte genomische und transkriptomische Analysen über die drei Stammzelllinien des Organismus hinweg durch. Diese Arbeit zeigte, dass verschiedene TE-Familien in jeder Zelllinie sowohl auf genomischer als auch transkriptioneller Ebene aktiv sind, und zwar durch nicht-zufällige Insertionen. Ein tief konservierter Kern von DNA-Transposons wurde als urtümlich aktiver Bestandteil tierischer Chromosomen identifiziert, der zu allmählichen genomischen Expansionen über Metazoen hinweg beiträgt. Diese Elemente beeinflussen strukturelle Variationen in der Nähe von Loci, die mit Zellproliferation und Chromatintopologie assoziiert sind, was darauf hinweist, dass urtümlich aktive DNA-Transposons als sich selbst erneuernde Bestandteile der Chromosomenarchitektur wirken.

In der zweiten Studie (Kon-Nanjo et al., *G3: Genes, Genomes, Genetics*, 2023) erstellte ich die erste chromosomenbasierte Genomassemblierung des kolonialen Hydrozoons *Hydractinia symbiolongicarpus* unter Verwendung von PacBio-HiFi-Langreads und Hi-C-Scaffolding. Die 483 Mb große Assemblierung, organisiert in 15 Chromosomen, die 99,8 % des Genoms abdecken, zeigte, dass 61 % aus repetitiven Sequenzen bestehen, was auf mindestens zwei Phasen der Wiederholungsexpansion hinweist. Mit 25.825 proteinkodierenden Genen (93,1 % BUSCO-Vollständigkeit) und einer ausgeprägten Makrosyntenie-Konservierung mit *Hydra vulgaris* stellt dieses Genom eine umfassende Grundlage für vergleichende Analysen der Hydrozoen-Stammzellbiologie und der chromosomalen Evolution dar.

In der dritten Studie (Kon-Nanjo et al., *Mobile DNA*, 2025) entdeckte ich eine massive subtelomerische Anreicherung einer einzelnen 372 bp langen Wiederholungseinheit, die mit der Helitron-DNA-Transposon-Familie verwandt ist, in *Hydractinia symbiolongicarpus*. Zusammen mit einer tandemartig angeordneten Minisatelliten-Wiederholung macht diese Elementfamilie 26 % des gesamten Genoms aus und erklärt weitgehend dessen Expansion im Vergleich zu anderen Nesseltiere. Vergleichende Analysen zeigten, dass Helitron-abgeleitete Sequenzen auch in der Nähe der Chromosomenenden anderer *Cnidaria* akkumulieren, was

darauf hindeutet, dass eine lokalisierte Amplifikation dieser Elemente subtelomerische Erweiterungen und stammespezifische Genomexpansionen innerhalb des Stammes antreiben kann.

Insgesamt zeigen diese Studien, dass transposable Elemente sowohl konservierte als auch dynamische Bestandteile tierischer Chromosomen sind. Sie formen die Diversifizierung der Genome, während sie gleichzeitig großräumige chromosomale Stabilität bewahren, und überbrücken damit die seit Langem bestehende Lücke zwischen chromosomaler Konservierung und evolutionärer Dynamik der Chromosomen.

Introduction

Deeply conserved synteny across animals

Animal genomes exhibit remarkable conservation at both chromosomal and sub-chromosomal levels (Simakov *et al.* 2022; Schultz *et al.* 2023). For example, a comparative genomic analysis across major metazoan (animal) lineages, including deuterostomes, protostomes, cnidarians, and sponges, revealed that specific sets of genes form ancestral linkage groups that have maintained their chromosomal associations since before the Cambrian period (Simakov *et al.* 2022). This finding indicates that the fundamental chromosomal architecture of animals was already established in early metazoan ancestors and has been stably preserved for more than 500 million years. Changes to this chromosomal complement were observed in sponges, ctenophores, other basally branching animal lineages, as well as unicellular outgroups. These patterns of chromosomal fusions suggested that the common ancestor of eumetazoans (the clade of animals with true tissues) underwent a series of chromosome reorganization events after its divergence first from the ctenophore and then from the sponge lineages, supporting the hypothesis that ctenophores represent the earliest branching animal group (Schultz *et al.* 2023). While such large-scale stability is often attributed to the constraints imposed by meiotic pairing and the deleterious effects of inter-chromosomal translocations in the germline (Wilch and Morton 2018), the evolutionary mechanisms preserving regulatory architectures within chromosomes remain less clear.

Despite the widespread conservation of chromosome homology, genome sizes differ substantially among species. The relative contributions of transposable elements (TEs) to these differences, however, are not yet fully understood. This raises an important question: are there any shared patterns underlying genome size variations, or is it largely stochastic? Although lineage-specific expansions of retroelements have been well documented (Nowoshilow *et al.* 2018; Meyer *et al.* 2021; Scharl *et al.* 2024), the roles of DNA transposons in shaping genome size evolution remain largely unexplored. Moreover, the temporal dynamics of TE bursts, their chromosomal localization patterns, and their long-term evolutionary fates can differ considerably even among closely related lineages (Chapman *et al.* 2010; Wong *et al.* 2019; Kon *et al.* 2025b). Understanding these processes is essential to clarify whether genome size evolution follows common mechanistic principles across metazoans or represents a lineage-specific, context-dependent outcome of genome-TE coevolution.

Genome evolution facilitated by TEs

TEs, also known as “jumping genes,” are DNA sequences that can move to different locations within a genome. TEs were first identified by Barbara McClintock in the 1940 through her studies of corn genetics (McClintock 1950). She demonstrated that certain DNA sequences are capable of moving from one chromosomal location to another, thereby influencing genetic variation and gene expression. Subsequent research revealed that TEs constitute a substantial portion of both plant and animal genomes (Feschotte and Pritham 2007). Advances in genome sequencing across a wide range of eukaryote taxa have revealed that the abundance and composition of TEs differ substantially among species (Lander *et al.* 2001; Chapman *et al.* 2010). For example, lungfish genome possesses more than ten times larger than that of humans, with TE content reaching up to 90% in some lungfish species (Scharl *et al.* 2024). In contrast, the pufferfish *Fugu rubripes* has one of the most compact vertebrate genomes, approximately 365 Mb in size, with TEs accounting for less than 3% of its total sequence (Aparicio *et al.* 2002). This extraordinary variation highlights the diverse

evolutionary histories and genomic strategies among eukaryotes regarding TE activity, retention, and elimination.

The fundamental classification of eukaryotic TEs, originally proposed by David Finnegan in 1989 (Finnegan 1989), divides them into two major classes based on their transposition intermediates: Class I (retrotransposons) and Class II (DNA transposons). Class I elements transpose through an RNA intermediate that is reverse-transcribed into DNA before reintegration into the genome (Wells and Feschotte 2020). However, with recent advances in the quantitative and qualitative characterization of TEs and their subfamilies, it has become clear that this traditional two-class framework no longer fully captures their diversity (Wells and Feschotte 2020). Among the most notable examples are Helitrons, a unique subclass of DNA transposons that replicate via a rolling-circle mechanism that typically leaves the donor site intact but can also involve partial excision or rearrangement events, unlike the canonical cut-and-paste pathway of Class II elements (Barro-Trastoy and Köhler 2024). Moreover, they lack terminal inverted repeats and target site duplications, two key hallmarks of typical DNA transposons (Barro-Trastoy and Köhler 2024). They frequently capture host gene fragments, thereby contributing to genome innovation and structural variation across animals and plants (Kapitonov and Jurka 2001, 2007).

In animals, Class I TEs, which propagate through a copy-and-paste mechanism, have been recognized as major contributors to several documented cases of genome expansion (Feschotte and Pritham 2007; Nowoshilow *et al.* 2018; Meyer *et al.* 2021; Scharl *et al.* 2024). In contrast, the contribution of Class II elements, which transpose through a cut-and-paste mechanism, remains less well understood, and their overall impact on genome size evolution is still unclear (Feschotte and Pritham 2007). Recent studies have also reported a decoupling between the accumulation of TEs and chromosomal conservation, as even extremely large animal genomes often maintain conserved synteny with few large-scale rearrangements (Meyer *et al.* 2021).

TEs and evolutionary novelty

Despite the deep conservation of genome architecture across animals (Simakov *et al.* 2022), gene expression patterns often diverge markedly among species, tissues, and cell types. TEs could be one of major contributors to this diversification of gene regulation (Feschotte 2008). By inserting into or near genes, TEs can supply new promoters, enhancers, and transcription factor binding sites, or alternatively disrupt existing regulatory elements, thereby altering expression levels and timing (Feschotte 2008). In some case, the same TE family disperses identical regulatory motifs across the genome, recruiting multiple genes into shared transcriptional networks (Peaston *et al.* 2004; Johnson *et al.* 2006). TEs can also affect post-transcriptional regulation by introducing alternative polyadenylation signals or microRNA (miRNA) target sites, and, in some instances, TE-derived sequences themselves evolve into new miRNAs (Smalheiser and Torvik 2006; Piriyaongsa *et al.* 2007). Furthermore, domesticated transposases, such as SETMAR in human (Cordaux *et al.* 2006) and FHY3 in *Arabidopsis* (Lin *et al.* 2007), have acquired gene regulatory functions. Through these diverse mechanisms, TEs serve as potent sources of evolutionary novelty, continuously reshaping gene regulation and contributing to the emergence of lineage-specific regulatory complexity in eukaryotes. It is also important to note that the environments surrounding the host organisms can be a major factor determining the evolution of TEs within their host genomes (Kofler *et al.* 2012, 2015, 2018).

Broader implications of TE research

TEs are not merely passive genomic passengers but dynamic agents that influence genome integrity, gene regulation, and evolutionary adaptation (Feschotte 2008). In the human genome, retrotransposons such as LINE-1, Alu, and SVA remain transcriptionally competent and can occasionally mobilize, generating structural variants that disrupt or rewire gene expression networks (Payer and Burns 2019). These insertions may alter transcription, splicing, or chromatin state, thereby contributing to phenotypic diversity and, in certain contexts, disease. For example, aberrant activation of LINE-1 has been implicated in tumorigenesis and age-related genomic instability (Xiao-Jie *et al.* 2016), while fixed endogenous retroviruses have been co-opted into regulatory roles essential for mammalian development, such as placental gene expressions (Chuong *et al.* 2013).

The dual nature of TE activity, with its potential to both disrupt the host physiological function and promote evolutionary innovation, highlights a broader evolutionary principle: mechanisms that pose risks to genome integrity can also act as powerful drivers of genetic novelty. Understanding how host genomes regulate and balance these opposing forces remains a central challenge in evolutionary genomics. Exploring TEs across a broader phylogenetic context, particularly in basally branching metazoans such as cnidarians, may uncover ancestral strategies of TE regulation and co-option that have contributed to the genomic plasticity observed across the animal kingdom (Kon-Nanjo *et al.* 2025).

Cnidarians represent an early-diverging metazoan phylum characterized by remarkable morphological and genomic diversity, encompassing solitary and colonial forms and exhibiting a wide range of life-history strategies (Daly *et al.* 2007; Steele *et al.* 2011; Technau and Steele 2011). Non-hydrozoan cnidarians, such as the anthozoan *Nematostella vectensis*, have become important model systems for studying the evolution of development and gene regulation due to their well-annotated genomes (Fletcher and Pereira da Conceicao 2023; Zimmermann *et al.* 2023) and experimental accessibility such as the CRISPR/Cas9-mediated genome editing and transgenesis (Rentzsch and Technau 2016). However, hydrozoans stand out within the Cnidaria because of their exceptional regenerative capacity, extensive diversity in stem-cell systems, and remarkable plasticity in colony-level organization (Tanaka and Reddien 2011; Plickert *et al.* 2012). Among them, *Hydra* and *Hydractinia* are particularly compelling comparative models: both possess experimentally tractable stem-cell lineages with clearly defined identities, yet they differ fundamentally in body organization and developmental potential (Plickert *et al.* 2012; Frank *et al.* 2020). Recent genomic studies have further revealed that their genomes have undergone lineage-specific TE-driven expansions, offering a unique opportunity to explore how TE dynamics intersect with stem-cell function and genome evolution within a single cnidarian class (Kon *et al.* 2025b; Kon-Nanjo *et al.* 2025).

Hydra as a model system for stem cell biology

Hydra is a freshwater cnidarian that represents the sister group to bilaterians, including vertebrates and insects. Cnidaria is one of the earliest diverging lineages of eumetazoans, a clade that comprises animals with true tissues, nervous and digestive systems, and a gastrula stage while excluding sponges (Kelava *et al.* 2015). Therefore, as a member of Cnidaria, *Hydra* occupies a pivotal phylogenetic position for understanding the evolution of multicellularity and complex body plans. It has also played a central role in the history of developmental biology, as both animal regeneration and the concept of the organizer were first discovered in this organism (Trembley 1744; Browne 1909). *Hydra* possesses a simple but highly dynamic body plan with a tubular structure composed of two epithelial layers, the

ectoderm and endoderm, separated by an extracellular matrix called the mesoglea, which functions similarly to the basement membrane in bilaterians (Reddy *et al.* 2019).

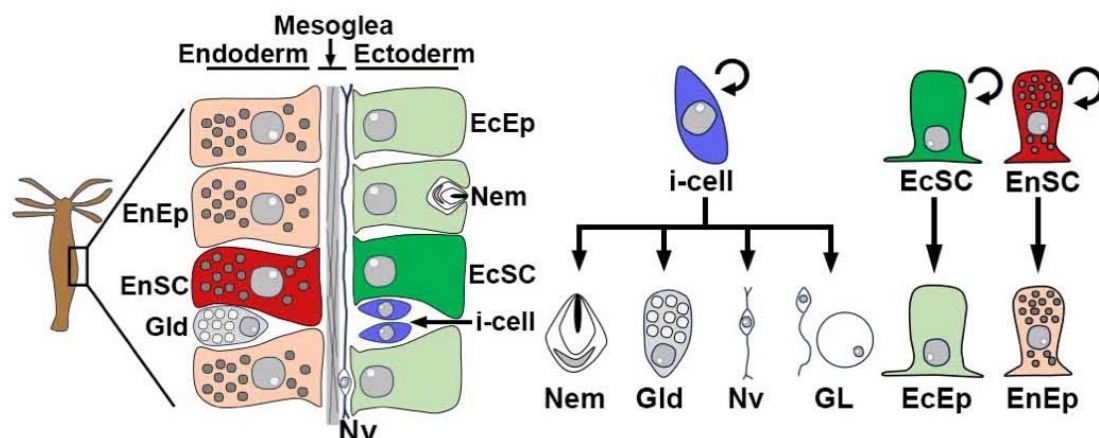


Figure 1. Hydra body plan and three stem cell lineages

This figure is a schematic representation of the cellular organization and stem cell lineages in Hydra. The body wall consists of two epithelial layers, the ectoderm (green) and endoderm (red), separated by the extracellular matrix (mesoglea). Three types of stem cell populations maintain tissue homeostasis. I-cells give rise to multiple somatic and germline cell types, including nematocytes (Nem), gland cells (Gld), nerve cells (Nv), and germline cells (GL). In contrast, ectodermal stem cell (EcSC) and endodermal stem cell (EnSC) differentiate into their respective epithelial cell types, ectodermal epithelial cells (EcEp) and endodermal epithelial cells (EnEp). Figure adopted from (Kon-Nanjo *et al.* 2025).

Hydra exhibits extraordinary regenerative capacity and biological immortality, showing no evidence of senescence (Martínez 1998; Schaible *et al.* 2015). Its ability to regenerate an entire individual from small tissue fragments or even from dissociated cells has made it a classical model for studying morphallactic regeneration and organizer-based patterning (Reddy *et al.* 2019; Vogg *et al.* 2019). The oral-aboral axis is continuously re-established through the coordinated action of conserved developmental signaling pathways such as Wnt, which regulate axis formation and tissue patterning (Hobmayer *et al.* 2000). Hydra reproduces both sexually and asexually through budding, allowing continuous somatic self-renewal and clonal propagation over extended generations (Vogg *et al.* 2019).

In addition to its anatomical and developmental simplicity, Hydra contains three distinct stem-cell lineages: interstitial stem cells (i-cells), ectodermal stem cells, and endodermal stem cells (Vogg *et al.* 2019) (Fig. 1). These cell types are responsible for both regeneration and tissue homeostasis. I-cells are multipotent and give rise to nerve cells, gland cells, nematocytes, and germline cells, whereas epithelial stem cells maintain the structural integrity of the ectodermal and endodermal layers. This continuous cellular turnover and clear lineage hierarchy provide an excellent framework to study the molecular principles of stem-cell maintenance and differentiation *in vivo* (Tanaka and Reddien 2011).

The establishment of transgenic Hydra lines expressing fluorescent markers in specific cell types, together with gene knockdown and genome-editing approaches such as RNA interference and CRISPR-Cas9, has greatly expanded its experimental potential (Nishimiya-Fujisawa and Kobayashi 2012; Lommel *et al.* 2017). The availability of a high-quality reference genome further facilitates molecular and genomic studies (Chapman *et al.* 2010; Simakov *et al.* 2022; Cazet *et al.* 2023), while recent single-cell RNA sequencing has

provided comprehensive transcriptomic profiles covering all major cell types across the entire organism (Siebert *et al.* 2019). These advances make Hydra an exceptionally powerful model for studying genome evolution, stem-cell biology, and cellular differentiation at both molecular and systems levels (Hobmayer *et al.* 2012).

Moreover, genomic studies have shown that TE activity is a major driver of genome expansion and genomic diversification in Hydra (Chapman *et al.* 2010; Wong *et al.* 2019), also having a potential role during regeneration (Petersen *et al.* 2015), suggesting that the interplay between TE dynamics, stem-cell function, and regeneration may hold the key to understanding how genome evolution contributes to long-term organismal plasticity and immortality.

Hydra TE-driven genome expansion and its impact on development and stem cell biology

Comparative genomic analyses across cnidarians have revealed striking differences in genome size within the genus Hydra. In most cnidaria such as green hydra (*H. viridissima*) and *Nematostella*, genome sizes are around 0.3 Gb (Hamada *et al.* 2020; Fletcher and Pereira da Conceicao 2023). In contrast, the brown hydras, including *H. vulgaris* and *H. oligactis*, possess genomes that are nearly three times larger, ranging from 1.0 to 1.5 Gb (Chapman *et al.* 2010) (Fig. 2). This expansion partially has been attributed to a lineage-specific proliferation of TEs, particularly the CR1 family of LINE retrotransposons, which underwent a massive burst of activity after the divergence of the brown and green hydra lineages approximately 59 million years ago (Wong *et al.* 2019). TEs and other repetitive elements now constitute more than 60 percent of the brown hydra genome, indicating that TE expansion rather than genome duplication was the main driver of genome size increase in this lineage (Kon-Nanjo *et al.* 2025).

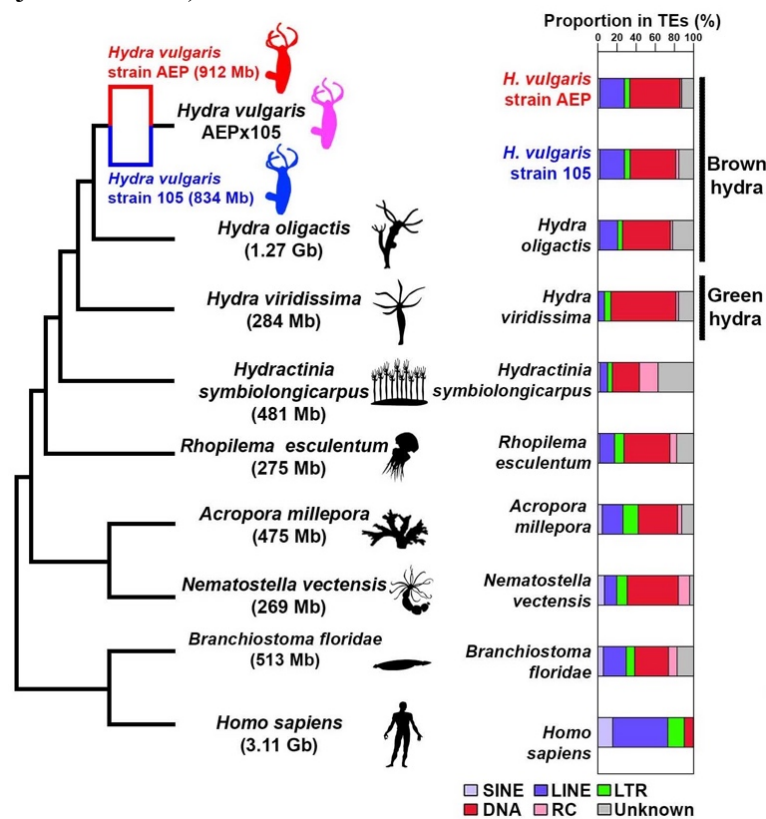


Figure 2. Genome size and TE composition across cnidarians

Phylogenetic relationships, estimated genome sizes, and the proportional composition of TE classes are shown for representative species. Brown hydras (*Hydra vulgaris* strains AEP and 105, and *H. oligactis*) possess substantially larger genomes (0.8–1.3 Gb) than the green hydra (*H. viridissima*, 0.28 Gb), reflecting lineage-specific genome expansion (Chapman *et al.* 2010; Wong *et al.* 2019). Figure adopted from (Kon-Nanjo *et al.* 2025).

Beyond their structural impact on hydra genomes, accumulating evidence suggests that TEs play active roles in the developmental and regenerative biology of Hydra. During head regeneration, several TE families show transient up-regulation within the first 30 minutes after bisection, particularly at the regenerating tip where organizer activity emerges (Petersen *et al.* 2015). Single-cell RNA sequencing has identified a distinct organizer cell cluster characterized by high TE transcription (Macias-Muñoz *et al.* 2022). The fact that TEs are transiently activated under regenerative stress conditions, together with observations that the highly regenerative axolotl genome is also enriched with TEs (Nowoshilow *et al.* 2018; Schloissnig *et al.* 2021), suggests that TEs may play conserved functional roles in regeneration and cellular stress responses.

The publicly available Hydra genome assemblies, while reaching chromosome scales, still contain numerous sequence gaps, and the subtelomeric regions including telomeric repeats are not yet completely assembled (Simakov *et al.* 2022; Cazet *et al.* 2023). Therefore, generating a more complete, telomere-to-telomere quality genome will be essential to uncover the detailed mechanisms and landscapes of TE-driven genome evolution in Hydra.

Hydractinia as a comparative hydrozoan model for stem cell evolution and comparative genomics

Hydractinia symbiolongicarpus is a colonial marine hydrozoan that inhabits gastropod shells occupied by hermit crabs along the Atlantic coast of North America (Frank *et al.* 2020). Colonies are composed of functionally specialized polyps for feeding, defense, and reproduction, which are interconnected by a stolonal network forming a shared gastrovascular system. The species can be readily cultured in the laboratory, where its predictable, light-induced spawning and rapid embryonic development have made it an attractive system for developmental and regenerative studies (Frank *et al.* 2020).

A defining feature of Hydractinia is its population of adult pluripotent stem cells, known as interstitial cells or i-cells, which can give rise to all somatic and germline lineages (Varley *et al.* 2023). Remarkably, a single transplanted i-cell is sufficient to regenerate an entire, sexually competent colony, demonstrating a degree of developmental potential not observed in Hydra. In contrast to the multipotent i-cells of *Hydra vulgaris*, which are restricted to the neuroglandular and germ cell lineages (Vogg *et al.* 2019), Hydractinia i-cells contribute to epithelial as well as germline and neural lineages, making this organism a unique model for studying the evolution of stem cell pluripotency among early-diverging animals.

Despite its biological importance, the genome of Hydractinia has not yet been fully resolved at the chromosome level. Existing assemblies remain fragmented (Schnitzler *et al.* 2024), limiting our ability to investigate global gene regulatory networks and chromosomal organization, including interactions between promoters, enhancers, and TEs (Kon-Nanjo *et al.* 2023). Establishing a chromosome-scale genome assembly will therefore be crucial to understanding the molecular basis of its regenerative and pluripotent capacities.

Furthermore, comparative genomics between *Hydractinia* and *Hydra* is expected to elucidate the evolutionary transitions in stem cell systems within hydrozoans. Such analyses will help identify the genomic and regulatory foundations that enabled the emergence of pluripotency in adult stem cells, providing key insights into the diversification of stem-cell strategies across early-branching metazoans.

Aim of the PhD project

Although TEs are now recognized as ubiquitous and evolutionarily significant components of eukaryotic genomes, our understanding of how they interact with genome architecture and contribute to regulatory and structural evolution remains fragmentary. Conventional classification frameworks and genome-scale analyses have largely treated TEs as passive sequences, while growing evidence indicates that they are dynamic, lineage-specific agents shaping genome diversification and functional plasticity.

To overcome these limitations, this PhD project aimed to elucidate the dynamics of active TEs in *Hydra vulgaris*, an organism renowned for its extraordinary regenerative capacity, asexual reproduction, and apparent biological immortality. In particular, I investigated which TE families are transcriptionally and genomically active in the three stem-cell lineages, including interstitial, ectodermal, and endodermal stem cells, and how their differential activity contributes to lineage-specific genome divergence during long-term clonal propagation. By integrating haplotype-resolved, telomere-to-telomere genome assemblies with stem-cell-specific genomic and transcriptomic data, this work sought to reveal how TE activity shapes chromosome architecture and long-term genomic plasticity in *Hydra* (Kon-Nanjo *et al.* 2025).

Building on these findings, I further aimed to extend the analysis to a broader comparative framework within Hydrozoa. Using *Hydractinia symbiolongicarpus*, a colonial cnidarian possessing adult pluripotent stem cells, I pursued the establishment of a chromosome-level genome assembly as a foundation for cross-species comparison. Comparative genomics between *Hydra* and *Hydractinia* is expected to illuminate the genomic and regulatory bases underlying the evolution of distinct stem-cell systems within cnidarians (Kon-Nanjo *et al.* 2023; Kon *et al.* 2025b).

Through these approaches, the overarching goal of this thesis was to uncover general principles of genome evolution driven by TE activity and to elucidate how deeply conserved chromosomal stability can coexist with the extraordinary genomic flexibility that characterizes early-branching metazoans.

Research approach overview

To address these aims, I first sought to overcome the limitations of existing TE analyses in the *hydra* genomes. The main challenges were threefold: (i) the ambiguity in distinguishing haplotypes within genome assemblies, (ii) the difficulty of reliably identifying TE insertion sites due to accumulated substitutions and indels, and (iii) the strong bias of previous studies toward intact, full-length elements with preserved coding regions, while fragmented or non-autonomous elements remained largely overlooked. These limitations have long hindered our ability to accurately reconstruct the full spectrum of TE diversity and their evolutionary impact on animal genomes.

To overcome these issues, I sequenced the AEP $\times$ 105 F1 hybrid genome, which enabled haplotype differentiation, and generated a haplotype-resolved, telomere-to-telomere (T2T)

assembly as a high-quality foundation for TE annotation. This assembly represents the first T2T genome in cnidarians and provides an essential resource for investigating TE-driven chromosome evolution. The AEP $\times$ 105 hybrid was produced from two genetically and geographically distinct Hydra strains, the North American strain AEP and the Eurasian strain 105 (Simakov *et al.* 2022). By analyzing their hybrid offspring, it became possible to distinguish paternal and maternal haplotypes and directly compare TE landscapes between them. This approach provided a unique opportunity to investigate how lineage-specific TE insertions contribute to chromosomal variation, genome expansion, and evolutionary divergence. Based on this genomic framework, I developed a pipeline to classify TEs into families and subfamilies, assess their genomic distribution, and quantify their contribution to large-scale chromosome architecture.

To refine TE detection, I systematically evaluated and optimized multiple annotation tools, including RepeatMasker and RepeatModeler. Special attention was given to enhancing both sensitivity and specificity by tuning parameters and implementing Hidden Markov Model (HMM)-based strategies for accurate identification of TE families and subfamilies. These efforts allowed a more detailed and unbiased reconstruction of TE landscapes across the Hydra genome.

In parallel, I established experimental resources to investigate TE activity at the level of stem-cell lineages. Transgenic Hydra lines were clonally expanded to minimize genetic drift, and large-scale cultures were maintained to capture potential lineage-specific TE accumulation during long-term propagation. Optimized cell dissociation and fluorescence-activated cell sorting (FACS) protocols enabled the isolation of high-quality genomic DNA from distinct stem-cell populations. The isolated high molecular genomic DNA from each stem-cell population was sequenced by the Nanopore long-read sequencing technology. These experiments provided the basis for examining how TE activity contributes to genome dynamics and chromosomal evolution in a stem-cell-resolved manner.

To place Hydra-specific findings within a broader evolutionary context, I applied the refined TE detection pipeline to the genomes of 82 representative eukaryotes. This comparative analysis aimed to identify conserved TE families and assess whether the patterns of TE dynamics observed in Hydra are shared across lineages. The results revealed the existence of an evolutionarily conserved core set of TE families among eukaryote genomes, suggesting that ancient, persistent TE groups, mostly DNA elements, have shaped chromosomal organization throughout evolution.

Finally, I extended these analyses to another hydrozoan model, *Hydractinia symbiolongicarpus*, which serves as an evolutionary bridge between Hydra and other metazoans with respect to stem-cell biology and genome organization. Hydractinia is a colonial cnidarian that possesses adult pluripotent stem cells (i-cells) capable of generating all somatic and germline lineages, making it an ideal system for comparative studies of stem-cell evolution. To establish a foundation for such analyses, I generated a high-quality, chromosome-level genome assembly of Hydractinia (Kon-Nanjo *et al.* 2023). This genomic resource enables detailed investigation of genomic TE composition and chromosomal architecture across cnidarian lineages. Through this analysis, I identified an extensive subtelomeric accumulation of Helitron-like DNA elements in Hydractinia chromosomes, suggesting a potential mechanism by which localized repeat amplification contributes to chromosomal extension and genome diversification in this species (Kon *et al.* 2025b).

My work and insights from Hydra on TE and chromosome biology and evolution has also contributed to several co-author publications during my thesis work (Ahuja *et al.* 2025; Shimada *et al.* 2025; Kon *et al.* 2025a; Yoshida *et al.* 2025).

References

- Ahuja, N., D. T. Schultz, D. Destanović, S. H. Church, N. Picciani *et al.*, 2025 A Chromosome-Scale Genome of *Nanomia septata* Reveals Extensive Rearrangement But No Clear Driver of the Unique Colony-Level Organization of Siphonophores. *bioRxiv*.
- Aparicio, S., J. Chapman, E. Stupka, N. Putnam, J. Chia *et al.*, 2002 Whole-Genome Shotgun Assembly and Analysis of the Genome of *Fugu rubripes*. *Science* (80-.). 297: 1301–1310.
- Barro-Trastoy, D., and C. Köhler, 2024 Helitrons: genomic parasites that generate developmental novelties. *Trends Genet.* 40: 437–448.
- Browne, E. N., 1909 The production of new hydranths in *Hydra* by the insertion of small grafts. *J. Exp. Zool.* 7: 1–23.
- Cazet, J. F., S. Siebert, H. Morris Little, P. Bertemes, A. S. Primack *et al.*, 2023 A chromosome-scale epigenetic map of the *Hydra* genome reveals conserved regulators of cell state. *Genome Res.*
- Chapman, J. A., E. F. Kirkness, O. Simakov, S. E. Hampson, T. Mitros *et al.*, 2010 The dynamic genome of *Hydra*. *Nature* 464: 592–6.
- Chuong, E. B., M. A. K. Rumi, M. J. Soares, and J. C. Baker, 2013 Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* 45: 325–329.
- Cordaux, R., S. Udit, M. A. Batzer, and C. Feschotte, 2006 Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc. Natl. Acad. Sci. U. S. A.* 103: 8101–6.
- Daly, M., M. R. Brugler, P. Cartwright, A. G. Collins, M. N. Dawson *et al.*, 2007 The phylum Cnidaria: A review of phylogenetic patterns and diversity 300 years after Linnaeus. *Zootaxa* 1668:.
- Feschotte, C., 2008 Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9: 397–405.
- Feschotte, C., and E. J. Pritham, 2007 DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41: 331–68.
- Finnegan, D. J., 1989 Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5: 103–107.
- Fletcher, C., and L. Pereira da Conceicao, 2023 The genome sequence of the starlet sea anemone, *Nematostella vectensis* (Stephenson, 1935). *Wellcome Open Res.* 8: 79.
- Frank, U., M. L. Nicotra, and C. E. Schnitzler, 2020 The colonial cnidarian *Hydractinia*. *EvoDevo* 11: 7.
- Hamada, M., N. Satoh, and K. Khalturin, 2020 A Reference Genome from the Symbiotic Hydrozoan, *Hydra viridissima*. *G3 (Bethesda)*. 10: 3883–3895.
- Hobmayer, B., M. Jenewein, D. Eder, M.-K. Eder, S. Glasauer *et al.*, 2012 Stemness in *Hydra* - a current perspective. *Int. J. Dev. Biol.* 56: 509–17.
- Hobmayer, B., F. Rentzsch, K. Kuhn, C. M. Happel, C. C. von Laue *et al.*, 2000 WNT signalling molecules act in axis formation in the diploblastic metazoan *Hydra*. *Nature* 407: 186–189.
- Johnson, R., R. J. Gamblin, L. Ooi, A. W. Bruce, I. J. Donaldson *et al.*, 2006 Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. *Nucleic Acids Res.* 34: 3862–77.
- Kapitonov, V. V., and J. Jurka, 2007 Helitrons on a roll: eukaryotic rolling-circle transposons.

- Trends Genet. 23: 521–9.
- Kapitonov, V. V., and J. Jurka, 2001 Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Sci.* 98: 8714–8719.
- Kelava, I., F. Rentzsch, and U. Technau, 2015 Evolution of eumetazoan nervous systems: insights from cnidarians. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:.
- Kofler, R., A. J. Betancourt, and C. Schlötterer, 2012 Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet.* 8: e1002487.
- Kofler, R., V. Nolte, and C. Schlötterer, 2015 Tempo and Mode of Transposable Element Activity in *Drosophila*. *PLoS Genet.* 11: e1005406.
- Kofler, R., K.-A. Senti, V. Nolte, R. Tobler, and C. Schlötterer, 2018 Molecular dissection of a natural transposable element invasion. *Genome Res.* 28: 824–835.
- Kon-Nanjo, K., T. Kon, H. R. Horkan, Febrimarsa, R. E. Steele *et al.*, 2023 Chromosome-level genome assembly of *Hydractinia symbiolongicarpus*. *G3 (Bethesda)*. 13:.
- Kon-Nanjo, K., T. Kon, T. C.-T. K. Yu, D. Rodriguez-Terrones, F. Falcon *et al.*, 2025 The dynamic genomes of *Hydra* and the anciently active repeat complement of animal chromosomes. *Genome Biol.* 26: 186.
- Kon, T., K. Kon-Nanjo, K. S. Handayani, L. Zang, F. Fahrurrozi *et al.*, 2025a Chromosome-level genome assembly of the doctor fish (*Garra rufa*). *Sci. data* 12: 765.
- Kon, T., K. Kon-Nanjo, and O. Simakov, 2025b Subtelomeric repeat expansion in *Hydractinia symbiolongicarpus* chromosomes. *Mob. DNA* 16: 14.
- Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody *et al.*, 2001 Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
- Lin, R., L. Ding, C. Casola, D. R. Ripoll, C. Feschotte *et al.*, 2007 Transposase-Derived Transcription Factors Regulate Light Signaling in *Arabidopsis*. *Science (80-.)*. 318: 1302–1305.
- Lommel, M., A. Tursch, L. Rustarazo-Calvo, B. Trageser, and T. W. Holstein, 2017 Genetic knockdown and knockout approaches in *Hydra*.
- Macias-Muñoz, A., L. Y. Mesrop, H. Y. Liang, and A. Mortazavi, 2022 Single cell analysis reveals the molecular signaling and cellular composition of the regenerating *Hydra* head. *bioRxiv*.
- Martínez, D. E., 1998 Mortality patterns suggest lack of senescence in *hydra*. *Exp. Gerontol.* 33: 217–25.
- McClintock, B., 1950 The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci.* 36: 344–355.
- Meyer, A., S. Schloissnig, P. Franchini, K. Du, J. M. Woltering *et al.*, 2021 Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* 590: 284–289.
- Nishimiya-Fujisawa, C., and S. Kobayashi, 2012 Germline stem cells and sex determination in *Hydra*. *Int. J. Dev. Biol.* 56: 499–508.
- Nowoshilow, S., S. Schloissnig, J.-F. Fei, A. Dahl, A. W. C. Pang *et al.*, 2018 The axolotl genome and the evolution of key tissue formation regulators. *Nature* 554: 50–55.
- Payer, L. M., and K. H. Burns, 2019 Transposable elements in human genetic disease. *Nat. Rev. Genet.* 20: 760–772.
- Peaston, A. E., A. V Evsikov, J. H. Graber, W. N. de Vries, A. E. Holbrook *et al.*, 2004 Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7: 597–606.
- Petersen, H. O., S. K. Höger, M. Looso, T. Lengfeld, A. Kuhn *et al.*, 2015 A Comprehensive Transcriptomic and Proteomic Analysis of *Hydra* Head Regeneration. *Mol. Biol. Evol.* 32: 1928–47.
- Piriyapongsa, J., L. Mariño-Ramírez, and I. K. Jordan, 2007 Origin and evolution of human

- microRNAs from transposable elements. *Genetics* 176: 1323–37.
- Plickert, G., U. Frank, and W. A. Müller, 2012 Hydractinia, a pioneering model for stem cell biology and reprogramming somatic cells to pluripotency. *Int. J. Dev. Biol.* 56: 519–34.
- Reddy, P. C., A. Gungi, and M. Unni, 2019 Cellular and Molecular Mechanisms of Hydra Regeneration, pp. 259–290 in.
- Rentzsch, F., and U. Technau, 2016 Genomics and development of *Nematostella vectensis* and other anthozoans. *Curr. Opin. Genet. Dev.* 39: 63–70.
- Schaible, R., A. Scheuerlein, M. J. Dańko, J. Gampe, D. E. Martínez *et al.*, 2015 Constant mortality and fertility over age in Hydra. *Proc. Natl. Acad. Sci.* 112: 15701–15706.
- Schartl, M., J. M. Woltering, I. Irisarri, K. Du, S. Kneitz *et al.*, 2024 The genomes of all lungfish inform on genome expansion and tetrapod evolution. *Nature* 634: 96–103.
- Schloissnig, S., A. Kawaguchi, S. Nowoshilow, F. Falcon, L. Otsuki *et al.*, 2021 The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. *Proc. Natl. Acad. Sci.* 118:.
- Schnitzler, C. E., E. S. Chang, J. Waletich, G. Quiroga-Artigas, W. Y. Wong *et al.*, 2024 The genome of the colonial hydroid Hydractinia reveals that their stem cells use a toolkit of evolutionarily shared genes with all animals. *Genome Res.*
- Schultz, D. T., S. H. D. Haddock, J. V Bredeson, R. E. Green, O. Simakov *et al.*, 2023 Ancient gene linkages support ctenophores as sister to other animals. *Nature* 618: 110–117.
- Shimada, Y., B. Aydın, K. Kon-Nanjo, K. S. Handayani, V. D. N. Gultom *et al.*, 2025 Potential of *Garra rufa* as a novel high-temperature resistant model fish: a review on current and future approaches. *Zool. Lett.* 11: 3.
- Siebert, S., J. A. Farrell, J. F. Cazet, Y. Abeykoon, A. S. Primack *et al.*, 2019 Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science* 365:.
- Simakov, O., J. Bredeson, K. Berkoff, F. Marletaz, T. Mitros *et al.*, 2022 Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* 8: eabi5884.
- Smalheiser, N. R., and V. I. Torvik, 2006 Alu elements within human mRNAs are probable microRNA targets. *Trends Genet.* 22: 532–6.
- Steele, R. E., C. N. David, and U. Technau, 2011 A genomic view of 500 million years of cnidarian evolution. *Trends Genet.* 27: 7–13.
- Tanaka, E. M., and P. W. Reddien, 2011 The cellular basis for animal regeneration. *Dev. Cell* 21: 172–85.
- Technau, U., and R. E. Steele, 2011 Evolutionary crossroads in developmental biology: Cnidaria. *Development* 138: 1447–58.
- Trembley, A., 1744 *Mémoires, pour servir à l’histoire d’un genre de polypes d’eau douce, à bras en forme de cornes*. Jean and Herman Verbeek, Leiden.
- Varley, Á., H. R. Horkan, E. T. McMahon, G. Krasovec, and U. Frank, 2023 Pluripotent, germ cell competent adult stem cells underlie cnidarian regenerative ability and clonal growth. *Curr. Biol.*
- Vogg, M. C., B. Galliot, and C. D. Tsiairis, 2019 Model systems for regeneration: Hydra. *Development* 146:.
- Wells, J. N., and C. Feschotte, 2020 A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* 54: 539–561.
- Wilch, E. S., and C. C. Morton, 2018 Historical and Clinical Perspectives on Chromosomal Translocations. *Adv. Exp. Med. Biol.* 1044: 1–14.
- Wong, W. Y., O. Simakov, D. M. Bridge, P. Cartwright, A. J. Bellantuono *et al.*, 2019 Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus Hydra. *Proc. Natl. Acad. Sci. U. S. A.* 116: 22915–22917.

- Xiao-Jie, L., X. Hui-Ying, X. Qi, X. Jiang, and M. Shi-Jie, 2016 LINE-1 in cancer: multifaceted functions and potential clinical implications. *Genet. Med.* 18: 431–439.
- Yoshida, M., E. Tóth, K. Kon-Nanjo, T. Kon, K. Hirota *et al.*, 2025 Giant genome of the vampire squid reveals the derived state of modern octopod karyotypes. *iScience* 113832.
- Zimmermann, B., J. D. Montenegro, S. M. C. Robb, W. J. Fropf, L. Weilguny *et al.*, 2023 Topological structures and syntenic conservation in sea anemone genomes. *Nat. Commun.* 14: 8270.

Results

Paper I

Title: “The dynamic genomes of Hydra and the anciently active repeat complement of animal chromosomes”

Authors:

Koto Kon-Nanjo^{1#}, Tetsuo Kon^{1*#}, Tracy Chih-Ting Koubkova Yu², Diego Rodriguez-Terrones³, Francisco Falcon^{3,4}, Daniel E. Martínez⁵, Robert E. Steele⁶, Elly Margaret Tanaka³, Thomas W. Holstein², Oleg Simakov^{1*}

¹ Department of Neurosciences and Developmental Biology, University of Vienna, Vienna, 1030, Austria

² Molecular Evolution & Genomics, Centre for Organismal Studies, Heidelberg University, Heidelberg, 69120, Germany

³ Institute of Molecular Biotechnology, Vienna, 1030, Austria

⁴ Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna, 1030, Austria

⁵ Department of Biology, Pomona College, Claremont, CA, 91711, USA

⁶ Department of Biological Chemistry, University of California, Irvine, Irvine, CA 92697-1700, USA

*Correspondence to Oleg Simakov (oleg.simakov@univie.ac.at), and Tetsuo Kon (kontetsuo@gmail.com); #Equally contributed

Status:

Published in Genome Biology, 2025, 26(1):186.

doi: 10.1186/s13059-025-03653-z.

RESEARCH

Open Access



The dynamic genomes of *Hydra* and the anciently active repeat complement of animal chromosomes

Koto Kon-Nanjo^{1†}, Tetsuo Kon^{1*†}, Tracy Chih-Ting Koubkova Yu², Diego Rodriguez-Terrones³, Francisco Falcon^{3,4}, Daniel E. Martínez⁵, Robert E. Steele⁶, Elly Margaret Tanaka³, Thomas W. Holstein² and Oleg Simakov^{1*}

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

*Correspondence:
kontetsuo@gmail.com;
oleg.simakov@univie.ac.at

¹ Department of Neurosciences and Developmental Biology, University of Vienna, Vienna 1030, Austria

² Molecular Evolution & Genomics, Centre for Organismal Studies, Heidelberg University, Heidelberg 69120, Germany

³ Institute of Molecular Biotechnology, Vienna 1030, Austria

⁴ Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna 1030, Austria

⁵ Department of Biology, Pomona College, Claremont, CA 91711, USA

⁶ Department of Biological Chemistry, University of California, Irvine, Irvine, CA 92697-1700, USA

Abstract

Background: Many metazoan genomes are characterized by highly conserved chromosomal homologies that predate the ancient origin of this clade. This conservation has been tested by expansions of selfish DNA elements, in particular transposable elements (TEs). While comparative genomics studies have highlighted their diversity across animal genomes, common principles underlying their evolution along deeply conserved chromosomes have been elusive. A detailed mechanistic understanding from phylogenetically key and early branching animal species has been lacking.

Results: We present a comprehensive stem-cell resolved genomic and transcriptomic study of the freshwater cnidarian *Hydra*, an animal characterized by its high regenerative capacity, the ability to propagate clonally, and an apparent lack of aging. Using single-haplotype telomere-to-telomere genome assemblies of two recently diverged strains and utilizing unique features of hydra biology allowed us to sequence and compare the individual genomes of hydra's three stem cell lineages. We show that distinct TE families are active at both transcriptional and genomic levels via non-random insertions in each of these lineages. We show that the core set of these active TE families, primarily composed of DNA elements, is evolutionarily deeply conserved and contributes to consistent genomic expansions in metazoan lineages. These anciently active TEs differentially contribute to structural variants around loci associated with cell proliferation and long-range topological contacts. This is in strong contrast to the frequently observed and highly varied substantial genome expansions that often happen via retroelements.

Conclusions: Our study suggests an ancient and conserved role for these core TEs as self-renewing components of animal chromosomes.

Keywords: Telomere-to-telomere genome assembly, Anciently active transposable elements, Genome expansion, Stem-cell genomes, Hydra



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

Animal genomes have been shown to be remarkably conserved at both chromosomal [1, 2] and sub-chromosomal levels [3–5]. While this remarkable conservation can be explained by gamete balancing issues that can arise in the case of inter-chromosomal translocations in the germline [6], the evolutionary forces that keep regulatory landscapes intact within chromosomes are less well understood.

One such force is the continuous expansion and contraction of chromosomes through the action of transposable elements (TEs). How this dynamic chromosomal landscape may contribute to macroevolution, and the resulting evolutionary constraints on chromosomal elements themselves, is unclear. Many cases of TEs contributing to genome expansions, changes in expression levels, regulatory element evolution, and structural variations have been reported (Additional file 1: Supplementary Note 1 [1, 7–40]). In animals, class I (copy-and-paste propagating TEs) have been especially strong contributors in many reported genome expansions [18, 41] (Additional file 1: Supplementary Note 1). The role of class II (cut-and-paste DNA elements) has been elusive and their broad contribution to genomic expansion remains unclear [18]. Recent studies also highlight a decoupling between repetitive element accumulation and chromosomal conservation (synteny), with extremely large animal genomes showing conserved synteny and few rearrangements [42]. To find common evolutionary and functional principles behind these genomic trends requires telomere-to-telomere haplotype-resolved genome assemblies and a detailed understanding of TE activity and insertion dynamics for key early branching animal species. Very few such resources are available.

The freshwater cnidarian *Hydra* sp. is a powerful and unique system for addressing these questions. As a member of the sister phylum to bilaterians, hydra occupies an important evolutionary position, spanning the earliest animal origin and diversification at least more than 550 million years ago based on fossils and molecular evolutionary analysis [43, 44]. *Hydra* was the first animal in which whole body regeneration was demonstrated [45], and the first animal in which a developmental organizer was identified [46–48]. *Hydra* serves as a simple yet powerful model system, providing profound insights into stem cell dynamics [49]. Adult polyps of *H. vulgaris* do not show signs of ageing [50, 51]. Homeostasis and the extensive regeneration capacity of hydra are maintained by three stem cell lineages, endodermal epithelial, ectodermal epithelial, and interstitial/germline stem cells (i-cells). Endodermal and ectodermal epithelial stem cells are unipotent, only producing epithelial cells. I-cell stem cells are multipotent and produce nematocytes, nerve cells, gland cells, and germline cells [52], though the rate may depend on the developmental stage [53]. The three stem cell lineages maintain their distinct identities and do not interconvert. *Hydra* regeneration from bisected animals, small tissue fragments, and even aggregates of cells has been extensively studied since the first experiments by Abraham Trembley in the 1700s [45]. More recently, transcriptomic studies have revealed the diversity of the molecular pathways that are active during hydra regeneration [32]. Comprehensive profiling of gene expression in each cell type and stem cell differentiation trajectories have been reported for the whole polyp of hydra through single-cell RNA-seq (scRNA-seq) [54]. While conserved components such as the Wnt signaling pathway have been investigated in developmental

and regeneration processes in hydra, many species-specific genes and, particularly, TEs are expressed, with unclear functional or evolutionary impact [32, 55, 56].

The ability of hydra to propagate asexually has striking implications for its biology. If hydras are cultured by budding for multiple generations, mutations accumulate in each stem cell lineage independently [57]. Therefore, clonally propagated hydras are expected to have three distinct genomes (ectodermal, endodermal, and i-cell stem cell lineages). If TEs have diverging functions in each of the three genomes, we expect to find differences in their cell-type specific transcriptional activity and genomic insertion profiles. These profiles may suggest potential impacts on cellular function and adaptation.

The genus *Hydra* sp. is composed of two major evolutionary lineages, the symbiotic green hydra (*H. viridissima*) and the species-rich brown hydras (e.g., *H. vulgaris* and *H. oligactis*) [58]. These two lineages diverged 46–61 million years ago [58], yet in this relatively short evolutionary time-span, brown hydra genomes have experienced a tripling in genome size [36, 40] while still preserving the ancient animal chromosomal homologies [1]. The green hydra genome is similar in size to genomes of most other cnidarians (less than 500 Mbp) [59]; the brown hydras have some of the largest cnidarian genomes (around 1 Gbp or more), with over 60% of the total sequence comprised repetitive sequences, including TEs [36, 60].

Within the brown hydras, two *Hydra vulgaris* strains, AEP and 105, have been extensively utilized in molecular studies. The strain AEP, from North America [61, 62], routinely undergoes sexual reproduction in laboratory conditions and is widely used to create transgenic lines [63]. Conversely, the strain 105, which was collected in Japan (described as *Hydra magnipapillata* strain 105), has been maintained through asexual reproduction in the laboratory since 1973 [64]. Although it almost certainly engaged in sexual reproduction in its natural environment, the presence of sexually mature individuals in the laboratory is extremely rare.

Together, these properties make hydra a valuable model system to investigate ancient evolutionary trends of TE composition and function in the animal kingdom. To uncover the role of the evolving TE landscape in hydra stem cell genome(s) and its impact on shaping the otherwise preserved, yet expanded, ancient animal chromosomal elements, we present the haplotype-resolved telomere-to-telomere genome assemblies of *Hydra vulgaris* 105 and AEP strains and reveal different TE evolutionary expansions among the three stem cell lineages. Using fluorescence-activated cell sorting (FACS) followed by genome sequencing, we reveal distinct TE insertion profiles in the three stem cell lineages in AEP-background strains. We show evidence indicating how these TE families impact the genome and their neighboring genes. We provide direct observation that the three stem cell genomes of hydra undergo distinct and non-random changes and suggest this as a key structural role of TEs in the evolutionary adaptations of these stem cells. Finally, through multi-species comparisons, we find remarkably similar TE signatures across eukaryotes and metazoans, identifying core sets of a dozen TE elements, mostly comprised DNA transposons, that have consistently shaped chromosomal evolution for hundreds of millions of years.

Results

Haplotype-resolved telomere-to-telomere reference genome assemblies for 105 and AEP strains of *H. vulgaris*

While chromosomal-scale assemblies of *H. vulgaris* strain 105 and strain AEP have been published [1, 60], they lacked the telomere-to-telomere completeness and haplotype

resolution needed to study the action of TEs in a stem cell lineage resolved manner (Fig. 1a–c). We utilized newly generated chromosomal conformational capture data from a AEP $\times$ 105 F1 hybrid (Methods) and new long-read sequencing to generate telomere-to-telomere assemblies (Fig. 1a, b, Additional file 1: Fig. S1a–h, Additional file 2: Tables S1–S3). As the sequencing data came from a hybrid of two recently diverged hydra strains (Fig. 1a, Additional file 1: Fig. S1a), our data produced a single haplotype for each of the strains with telomere-to-telomere quality (Fig. 1a, b, Additional file 1: Figs. S1b–h, S2a, b). Both strains showed a high degree of local (micro-) synteny conservation and the previously reported conservation of ancient animal chromosomal elements (Fig. 1b) [1]. However, the 105 strain genome showed a single apomorphic translocation between chromosome 5 and chromosome 15, which was previously suggested [1] and can now be confirmed with our data (Fig. 1b). Our analysis further identifies this event as a Robertsonian translocation where chromosome arms are exchanged at the centromeric regions (Fig. 1b, Additional file 1: Fig. S2c–e).

When comparing the 13 chromosomes, excluding the chromosomes that underwent the Robertsonian translocation, between each haplotype genome, 12 chromosomes of the AEP haplotype were longer than those of the 105 haplotype suggesting that the AEP genome has experienced a genome-wide size increase (Fig. 1b, Additional file 1: Fig. S1e). We found that the major groups of TEs (including SINE, LINE, LTR, DNA, RC; Fig. 2a) accounted for 65.3% (596 Mb) of the AEP genome and 63.3% (528 Mb) of the 105 genome (Fig. 2b, c). Among these groups, DNA transposons were the most common,

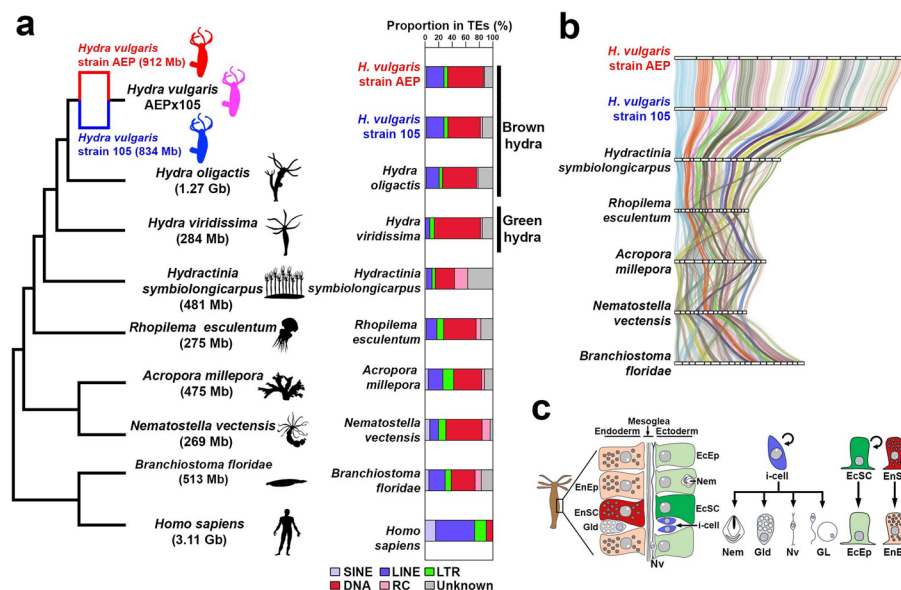


Fig. 1 Haplotype-resolved telomere-to-telomere genome assemblies for 105 and AEP strains of *H. vulgaris* highlighting genome expansion. **a** Cladogram representing major cnidarian species. Numbers in parentheses indicate the sizes of publicly available genome assemblies. Animal silhouettes are from PhyloPic (www.phylopic.org). The bar plot represents the proportions of TEs relative to the total repetitive elements in the various genomes. **b** Comparative genomic synteny. The coordinates of 2440 single-copy orthologs from 6 cnidarians and one bilaterian (*Branchiostoma floridae*) are connected by lines. **c** The stem-cell system of hydra. Abbreviations: i-cell, interstitial stem cells; EcSC, ectodermal stem cell; EnSC, endodermal stem cell; Nem, nematocyte; Gld, gland cell; Nv, nerve cell; GL, germline cell; EcEp, ectodermal epithelial cell; EnEp, endodermal epithelial cell

comprising 34.2% and 30.1%, followed by LINEs at 16.7% and 16.3% in the AEP and 105 genomes, respectively (Fig. 2b, c). When comparing the age distribution (Kimura substitution levels) among each major TE group between the AEP and the 105 haplotypes, we found that relatively young DNA elements and LINEs have increased in AEP, suggesting their contribution to the genome size increase after the divergence from the last common ancestor of AEP and 105 (Fig. 2d).

We used these genome assemblies to compare the two strains in detail and to dissect the repeat composition along the individual chromosomes. We note that automatic TE family annotation is prone to many genomic and bioinformatic artifacts [65]; therefore, we cross-validated our annotation using other available annotation databases, including Dfam [66], Repbase [67], and Gypsy Database [68] (Additional file 1: Supplementary Note 2 [67, 68]). While classification and the level of manual curation vary among these resources, we found generally consistent result of 106 uniquely assigned and annotated TE families present in the hydra genome (Additional file 2: Table S4), similar to the previously published estimates [36, 40]. The genome coverage difference of TEs between the

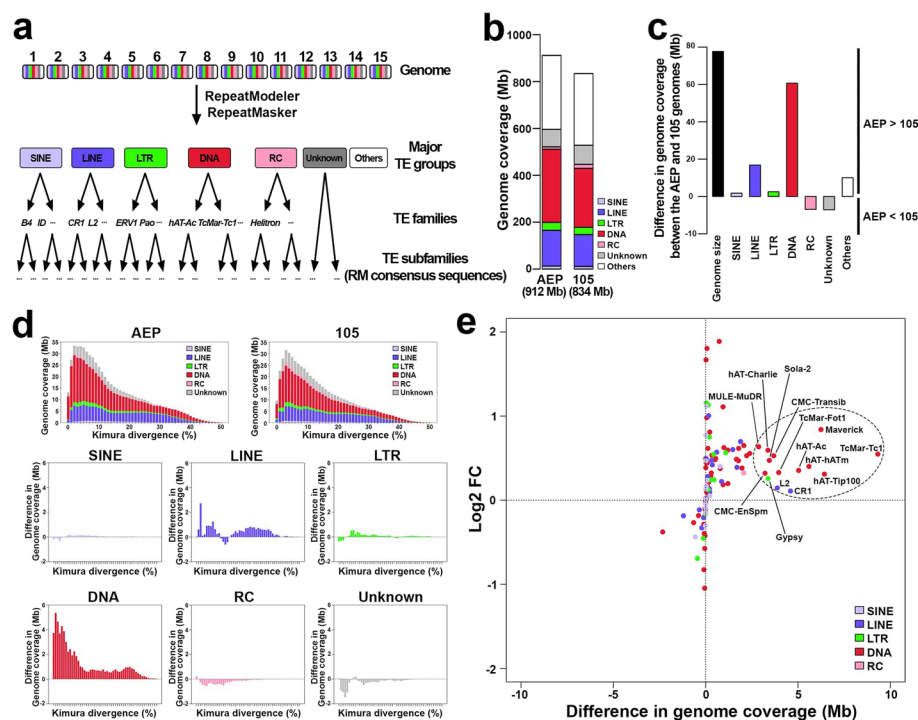


Fig. 2 DNA transposons as the major contributor behind the recent AEP genome expansion. **a** TE classification pipeline and levels of TE annotation. Abbreviation: RM, RepeatModeler. **b** TE genome coverage in the AEP and 105 genomes. **c** Difference in TE genome coverage between the AEP and 105 genomes. **d** Kimura divergence and TE genome coverage. The top two panels show Kimura divergence and TE genome coverage for the AEP and 105 genomes. The bottom six panels represent Kimura divergence and difference in TE genome coverage between the AEP and 105 genomes, showing enrichment and recent insertions of mostly DNA elements, along with LINE and LTR elements, in the AEP strain. The scale of the x-axis in the bottom six panels is the same as that in the top two panels. **e** Difference and log2 fold change in TE genome coverage between the AEP and 105 genomes. The x-axis represents the difference in genome coverage of TE families between the AEP and 105 genomes (in Mb), while the y-axis displays the log2-transformed fold change (Log2 FC) in genome coverage of TE families between the two genomes. The dotted ellipse highlights a cluster of 14 TE families with over 2.5 Mb greater genome coverage in the AEP genome compared to the 105 genome (A-TEs)

AEP and 105 haplotype genomes is 67.6 Mb, which accounts for 87.1% of the genome size difference between the two genomes (Fig. 2c).

Using this detailed annotation, we identified the putatively active TEs (“A-TEs”) composed of 14 TE families with the enrichment in AEP relative to 105 (Fig. 2e, Additional file 1: Fig. S3, Additional file 2: Table S4, Additional file 1: Supplementary Note 2). These included 11 DNA elements (hAT-Ac, hAT-Charlie, hAT-hATm, hAT-Tip100, TcMar-Tc1, TcMar-Fot1, Maverick, Sola-2, MULE-MuDR, CMC-EnSpm, and CMC-Transib), two LINE elements (CR1 and L2), and one LTR element (Gypsy superfamily). These elements were found to contribute the most to the AEP genome size increase (Fig. 2e, Additional file 2: Table S4). A-TEs accounted for 36.6% (334 Mb) of the AEP genome and 35.8% (299 Mb) of the 105 genome, while non-A-TEs accounted for 28.7% (262 Mb) of the AEP genome and 27.5% (229 Mb) of the 105 genome (Additional file 2: Table S4). Although the genome size increase attributable to the non-A-TEs (93 families, 1.2%) collectively was slightly higher than that of the A-TEs (14 families, 0.8%), the per-family contribution of each non-A-TE family was substantially smaller compared to that of the active TE families. While the 14 A-TEs collectively affect whole chromosomes, some subfamilies of A-TEs showed local propagation in specific chromosomal regions in AEP characterized by the lack of local synteny, suggesting expansion through tandem duplication (Additional file 1: Fig. S4a, b).

This analysis, based on complete and haplotype-resolved assemblies, highlights a diverse set of putatively active TE (A-TE) families in the hydra genome, which suggests a more complex picture of the brown hydra genome expansion than the previous finding of a single retroelement contribution to the genome size increase in this clade [36, 40]. We next sought to investigate these A-TEs transcriptionally.

Expression activity of transposable elements in hydra

While expression of TEs has been proposed in hydra in previous studies [32, 69], validation using different sequencing technologies has been lacking. To identify a high confidence set of expressed TEs, we analyzed an Iso-Seq transcriptome generated from whole polyps of the 105 × AEP hybrid line. We identified 29,833 transcripts (0.32%) derived from TEs containing at least 1 kb of annotated sequence, including 14,302 transcripts (0.15%) from the AEP haplotype genome and 15,531 transcripts (0.16%) from the 105 haplotype genome. These transcripts included the major groups of TEs (Fig. 3a, b) and encompassed 86 different TE families (75 families and 73 families for the AEP and 105 haplotype genomes, respectively) (Fig. 3c). The average transcript length suggests that near full-length transposon sequences are being captured by Iso-Seq, with LTRs exhibiting the highest average length of 3.5 kb in the 105 haplotype genome (Fig. 3a). We also find that 30.9% of TE Iso-Seq reads in AEP and 29.7% of TE Iso-Seq reads in the 105 strain have trans-spliced leader (SL) sequences [36, 70] at the 5′ end of the reads. For Iso-Seq reads of non-TEs, 48.1% were found to have SL sequences. These sequences are associated with a reaction known as spliced leader trans-splicing, in which SL RNAs are added to the 5′ end of mRNAs through a trans-splicing mechanism [71]. This suggests that transcripts of TEs are utilizing hydra’s endogenous trans-splicing mechanism (Additional file 1: Fig. S5a, b), similar to the protein-coding genes [36]. All previously identified putatively active TEs (A-TEs) showed expression (Fig. 3c, Additional file 1:

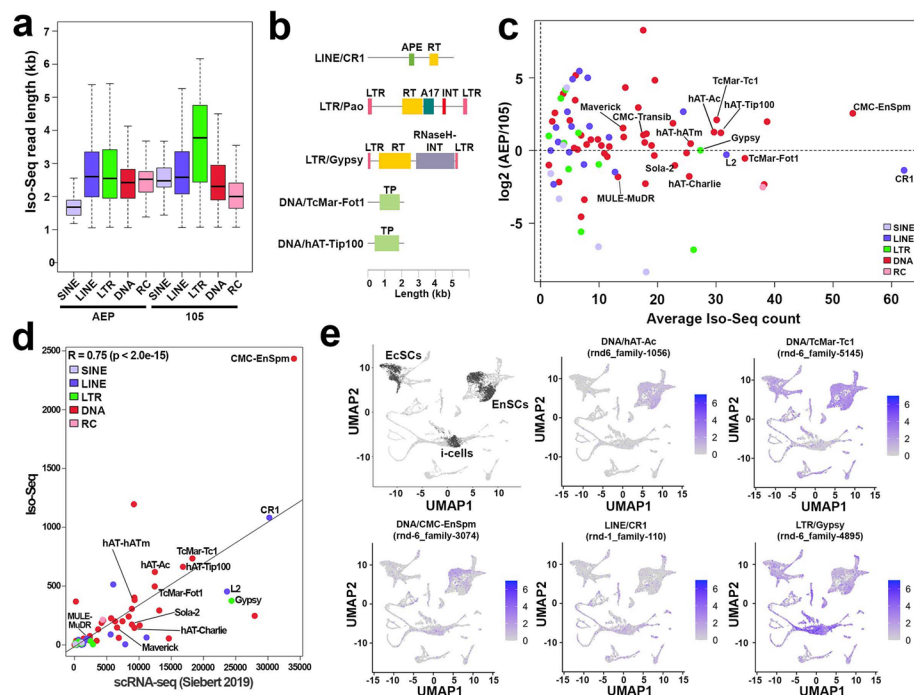


Fig. 3 Strain and cell-type specific expression dynamics of hydra TE families. **a** The lengths of Iso-Seq reads capturing TE sequences. The boundaries of the box depict the first (25%) and third (75%) quartiles, while the median is denoted by the central horizontal line within the box. The whiskers extending from the box indicate the lowest and highest observed values. **b** Schematic domain structures of TEs identified by the Iso-Seq analysis. Abbreviations: APE, apurinic endonuclease; LTR, long terminal repeat; RT, reverse transcriptase; A17, peptidase A17; INT, integrase; RNaseH, ribonuclease H; TP, transposase. **c** TE expression difference between AEP and 105 strain. The x-axis represents the average transposable element (TE) expression between the AEP and 105 strains, while the y-axis displays the Log2 fold change in TE expression between them. **d** Correlation between Iso-Seq data from the present study and publicly available data. The x-axis represents the average transposable element (TE) expression in publicly available scRNA-seq data [54], while the y-axis shows the average TE expression in the Iso-Seq data from the present study. **e** TE expression profile at the cell-type level. In the left upper panel, stem cell populations are colored in black and other differentiated cells colored in gray. The rest of panels show representative plots of A-TE subfamily expressions

Fig. S5c, d, Additional file 1: Supplementary Note 2). Of all A-TEs, 11 A-TEs, excluding CMC-Transib, Maverick, and MULE-MuDR, were among the top 20% of the 86 TE types according to the expression level (Fig. 3c). Thirty-three percent of TE Iso-Seq reads were mapped to intronic regions, 66% of TE Iso-Seq reads were mapped to intergenic regions, and less than 1% were mapped to exons, across over 4087 and 4672 loci in the 105 and AEP genomes, respectively (Additional file 1: Fig. S5c, e).

Next, we analyzed the expression profiles of TEs using the publicly available single-cell RNA-seq dataset from adult polyps of the AEP strain [54]. We quantified the expression levels of TEs that had been confirmed to be expressed by Iso-Seq in the whole polyp (Fig. 3a–d). A significant correlation was observed between the expression levels of TEs in Iso-Seq and scRNA-seq ($R=0.75$, $p<2.0e-15$; Fig. 3d). We investigated TEs that showed significantly higher expression levels in one or more of the six cell populations—comprising interstitial stem cells, cells differentiated from interstitial stem cells, ectodermal stem cells, cells differentiated from ectodermal stem cells, endodermal stem cells, and cells differentiated from endodermal stem cells—compared to the rest of the

cell populations (Fig. 3e, left upper panel). At the family level, six TEs (TcMar-Tc1, CMC-Chapaev-3, PIF-Harbinger, hAT-Charlie, hAT-hAT5, and MULE-NOF) showed specific expression in one or more cell populations, with TcMar-Tc1 and hAT-Charlie classified as A-TEs (Additional file 2: Table S5). All A-TEs were expressed at least in i-cells which is consistent with the fact that i-cells can give rise to germline cells [53], whose genomes are inherited by all three stem cell lineages through sexual reproduction (Additional file 1: Figs. S5d, S6, Additional file 2: Tables S6, S7, Additional file 1: Supplementary Note 2).

In summary, these data provide a comprehensive bulk, tissue, and cell-type resolved cross-validation analysis of hydra TE expression dynamics, revealing distinct and cell-type specific patterns of TE expression and the high expression activity of the A-TE families.

Identification of stem cell-specific transposon insertion sites using long-read sequencing

Hydra, primarily reproducing asexually through budding, can also resort to sexual reproduction. In laboratory conditions, transgenic hydra strains are propagated asexually, with offspring inheriting genomes from three parental stem cell types. Over time, these strains may accumulate unique transposon insertions in each stem cell type's genome (Additional file 1: Fig. S7a). Utilizing this aspect of hydra biology, we investigated the genomic diversity caused by TE insertions in stem cells using the Cnno1::GFP hydra transgenic line, which labels interstitial stem cells under the Nanos promoter (Fig. 4a) [53, 72]. A clonal population propagated through budding from a single founding polyp for more than 10 years was subjected to FACS to isolate GFP-positive (i-cells) and GFP-negative cells (non-i-cells), followed by Nanopore PromethION long-read sequencing for each group, yielding 15.1 Gb and 16.3 Gb of long reads for i-cells and non-i-cells, respectively (Fig. 4b, Additional file 1: Fig. S7b). These reads were aligned to the AEP genome sequence for variant calling with filtering (Additional file 1: Fig. S7b, Methods).

We identified 54,857 genomic insertions across both populations, with 6506 larger than 1 kb, indicative of active TEs (Fig. 4c, Additional file 2: Table S8). Notably, 726 insertions were unique to i-cells, 423 to non-i-cells, and 5357 were shared (Fig. 4d). The higher number of TE insertions in the i-cells correlates with the faster propagation rate of i-cells in homeostatic tissue [73, 74]. TEs constituted a major part (69.7%) of these insertions, with notable contributions from the A-TEs (Fig. 4e, f). Among the TE insertions, 68.0% (494 insertions) were in the i-cell population, 65.0% (275 insertions) in the non-i-cell population, and 70.3% (3767 insertions) shared by both populations, respectively (Fig. 4d). The 5357 insertions including the 3767 TE insertions which were observed both in i-cell population and non-i-cell population can be regarded as reflecting the genetic background difference between the transgenic line and the AEP line used for the reference genome (Fig. 4d). Among these TE insertions, DNA elements were the most numerous, followed by LINE elements, and then LTR elements (Fig. 4e). These TE insertions were composed of 55 TE families, 50 TE families, and 81 TE families for the i-cell population, non-i-cell population, and shared population, respectively (Fig. 4d). All of the populations included the A-TEs (Fig. 4f–h, Additional file 1: Fig. S7c, Additional file 1: Supplementary Note 2).

Using single-cell transcriptome and open chromatin data, we investigated the role of TE insertions in gene expression variation relative to insertion regions. We found that

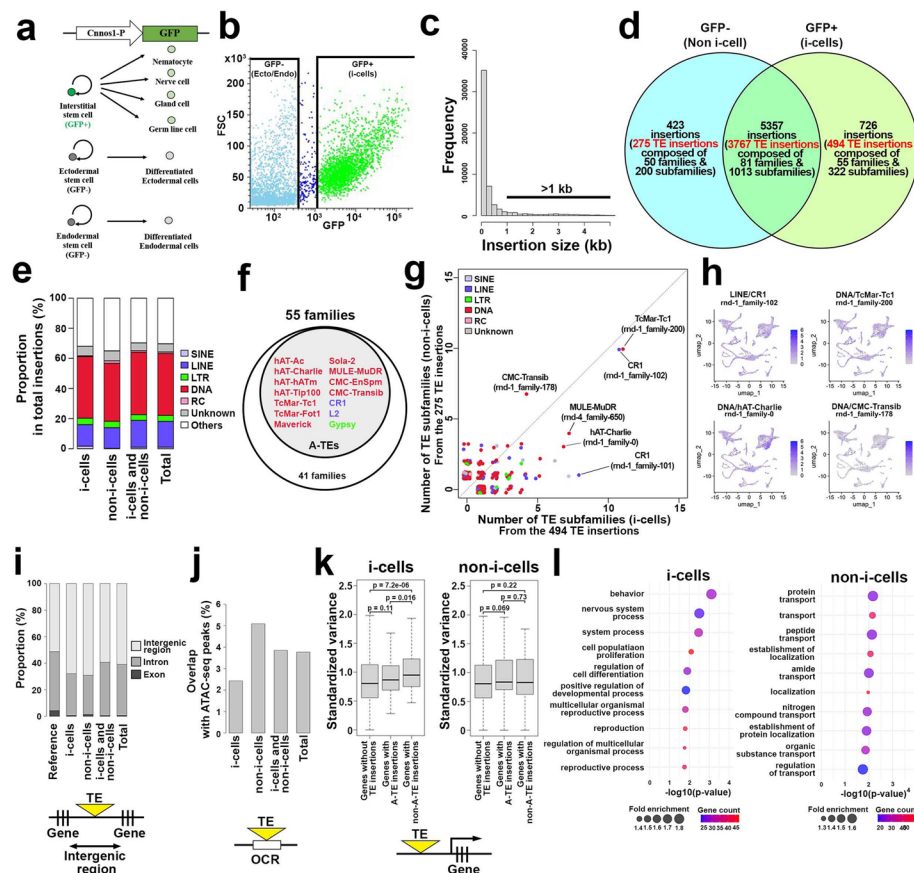


Fig. 4 TE insertion dynamics of stem cells in hydra. **a** Cnno1-GFP transgenic line labeling i-cell stem cells used for this analysis. **b** FACS-sorting of the i-cell population and non-i-cell population. **c** Size distribution of genomic insertions. **d** Number of cell type specific and shared insertions. Some of the insertions were further annotated as TEs highlighted in red. **e** Major types of TEs contributing to the genomic insertions. **f** TE families which comprise the genomic insertions of the i-cell population. The core 14 TE families are shown in the gray circle. **g** Counts of TE subfamilies in the 494 TE insertions (x-axis) and 275 TE insertions (y-axis). **h** Expression profile of TEs showing genomic insertions in the stem-cell lineage, based on reanalysis of publicly available scRNA-seq [54]. **i** Distribution of TE genomic insertions in intergenic regions, introns, and exons among cell populations. **j** Proportion of TE genomic insertions overlapping with open chromatin regions (OCRs). The bar plot shows the percentage of TE insertions that fall into OCRs, highlighting their potential regulatory significance. **k** Relationship between TE genomic insertions and expression variance. The y-axis represents the standardized expression variance of genes with or without nearby A-TE or non-A-TE genomic insertions in the present whole-genome sequencing of stem cells. The standardized expression variance of genes was calculated from the publicly available scRNA-seq data [54]. **l** GO-term enrichment analysis of closest genes to the insertions detected in the i-cell as well as non-i-cell (ecto/endo stem cells) lineage genomes

insertions generally happen more frequently in the intergenic regions for both i-cell and non-i-cell lineages, compared to the genomic average of fixed TE copies (Fig. 4i). However, striking differences were found between TE insertions in their open chromatin insertion preferences depending on stem cell lineage (Fig. 4j). TE insertions in the i-cell lineage occurred less frequently in open chromatin regions, with only 2.4% overlapping ATAC-seq peaks (p value = 0.002, Fisher's exact test; Fig. 4j). In contrast, the non-i-cell lineage exhibited a higher frequency of TE insertions overlapping with ATAC-seq peaks compared to other lineages (over 5% of TE insertions, Fig. 4j). Corroborating this finding, A-TEs contributed to less gene expression variation in the i-cells compared to the

other lineages (Fig. 4k). Our data suggest that open-chromatin TE insertions tend to be associated with higher expression variation in the somatic lineages. Furthermore, this suggests that TE insertions are under tighter regulatory control in the i-cells. The activity of TEs in hydra stem cells is known to be regulated by the PIWI-piRNA pathway [75, 76]. Consistent with this, scRNA-seq data shows that Piwi2 is expressed in three types of stem cells, with particularly high expression in i-cells (Additional file 1: Fig. S7d).

These data are supported by distinct categories of genes affected by stem-cell lineage-specific insertions. Analysis of the genes near the TE insertion loci specific to the i-cell population revealed 20 statistically significantly enriched gene ontology (GO) terms such as cell proliferation (left panel in Fig. 4l). This enrichment was highlighted by the i-cell specific A-TE hAT-Tip100 insertion in the intron of 3-phosphoinositide-dependent protein kinase 1 (PDPK1) [77, 78] (Fig. 4l, Additional file 1: Fig. S7e). This hAT-Tip100 genomic insertion contains a target site duplication, suggesting that this insertion occurred relatively recently (Additional file 1: Fig. S7f). Similarly, 22 gene groups associated with TE insertions specific to non-i-cells were identified, with GO terms related to substance transport such as protein transport and peptide transport being significantly enriched (right panel in Fig. 4l).

Similarly, we performed stem-cell lineage-specific whole-genome sequencing on the ecto-GFP/endo-RFP line [79], which expresses GFP in ectodermal epithelial stem cell lineage and RFP in endodermal epithelial stem cell lineage and has been propagated by budding since 2011 (Additional file 1: Fig. S8a–d, Additional file 1: Supplementary Note 2). We identified 409 TE insertions (54 families), 523 TE insertions (61 families), and 464 TE insertions (59 families) specific to the ectodermal stem cell lineage cell population, endodermal stem cell lineage cell population, and non-epithelial cell population, respectively (Additional file 1: Fig. S8c, d, Additional file 2: Table S9). Similar to the TE insertion patterns in the Cnno1::GFP line (Fig. 4f, g, Additional file 1: Fig. S7c), these insertions were predominantly from the A-TEs (Additional file 1: Fig. S8e–h). However, they were associated with genes belonging to different GO categories, such as cilium organization in ectoderm, chromosome organization in the endoderm, and monoatomic ion transport in the non-epithelial cells (Additional file 1: Fig. S8i), suggesting different targeting of A-TE in distinct stem cell lineage genomes.

In summary, our lineage-specific whole-genome sequencing of stem cells has unveiled loci for TE insertions associated with distinct gene ontologies and predominantly influenced by the A-TEs. In addition, the contribution of the A-TEs to the general transgenic line differences suggests their accumulation in the germline.

Deep evolutionary conservation of the A-TE families

We investigated whether A-TEs are present across a wide clade in the genomes of animals and their outgroups to test the phylogenetic scope of their putative activity beyond hydra. For this purpose, we developed a custom deep homology search, incorporating hidden Markov model-based sequence searches to generate a custom repeat library with TE annotation (Additional file 1: Fig. S9a, [Methods](#)). Our pipeline has substantially improved the detection of TEs in the genomes of three unicellular holozoans (*Creolimax fragrantissima*, *Capsaspora owczarzaki*, *Salpingoeca rosetta*) compared to the standard methodology (Additional file 1: Fig. S9b–d). We analyzed TE homology and retention in

82 eukaryotic species representing all major eukaryote clades [80, 81] (Fig. 5a and Additional file 1: Fig. S10, Additional file 2: Tables S10, S11). A total of 178 TE families were detected in at least one species analyzed (Fig. 5b and Additional file 1: Fig. S10). We discovered a cluster of several TE families with overrepresentation across all the species (cluster #1 in Fig. 5c and Additional file 1: Fig. S10). We found that at least 58.8% (ten out of 17 TE families) of these families are members of A-TEs (arrows in cluster #1 in Fig. 5c and Additional file 1: Fig. S10). Only species with very compact genomes lacked a clear A-TE signature (Fig. 5c and Additional file 1: Fig. S10). The pan-eukaryotic distribution of these TE families suggests a very ancient conservation and roles of these A-TEs in shaping chromosomes that might predate the current metazoan-unicellular homologies [2]. We term these families core eukaryotic TEs. We found that the other part of A-TEs are located within one of several pan-metazoan clusters (cluster #2 in Fig. 5c and Additional file 1: Fig. S10). While this cluster has a particularly high activity in the genus *Hydra*, these TE families are also abundant in other metazoan species and may thus represent an active core metazoan TE set.

We found evidence that core TEs are continuously shaping animal chromosomes. We profiled the reported genome expansion in the zebrafish genome [82] and found that the core eukaryotic TE cluster members are heavily involved in the *Danio rerio* genome expansion contributing to 30.0% of the genome (Additional file 2: Table S11). We found that the core metazoan TE set is present but not as prominently active in the zebrafish genome as in the brown hydra genomes (cluster #2 in Fig. 5c and Additional file 1: Fig. S10).

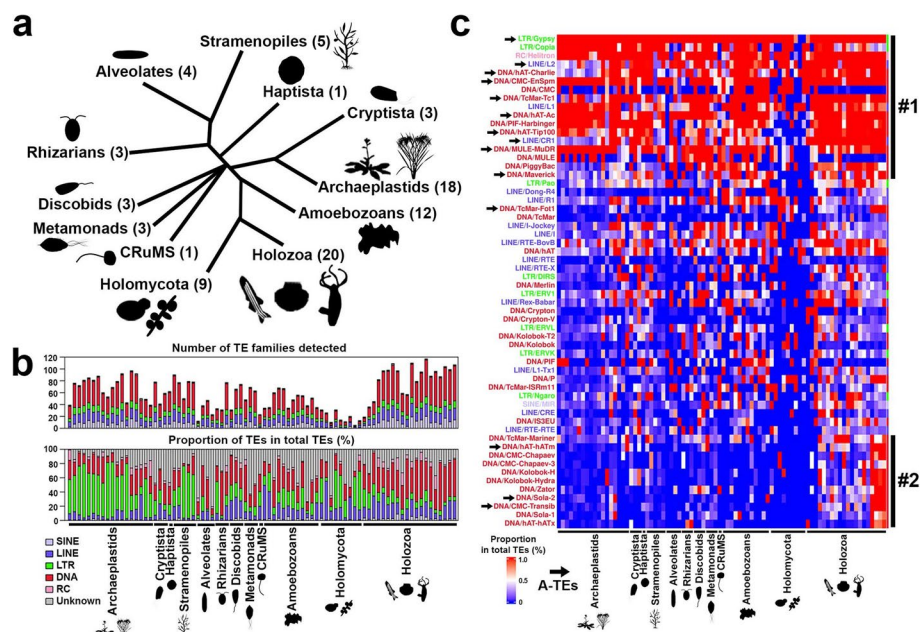


Fig. 5 Deep TE family homology and retention of genomically active TEs across eukaryotes. **a** The cladogram of the species used in this analysis. The topology of the cladogram is derived from previous research [80, 81]. **b** Number of TE families detected in each species (upper panel) and proportion of TEs in total TEs in each species (lower panel). **c** The heatmap shows the extent to which the genome coverage of each TE family constitutes the overall genome coverage of TEs in each species. TEs belonging to cluster #3 in Additional file 1: Fig. S10 are shown. The cluster #1 (eukaryotic core TEs) and #2 (active in *Hydra*) correspond to those in Additional file 1: Fig. S10. The 14 TEs of the A-TEs are indicated by arrows

Similarly, nine out of 17 core eukaryotic TEs have contributed to the ancient coleoid cephalopod genome expansion and evolution [83]. However, we also found other TE clusters that are expanded in species-specific manners, that are not covered by A-TEs and may thus constitute divergent lineage-specific active TE core sets that were not captured in our hydra-centric analysis, including several ERVs and L1 families (Fig. 5c and Additional file 1: Fig. S10, Additional file 2: Tables S12–S15, Additional file 1: Supplementary Note 2).

Taken together, our data suggest that A-TEs are forming evolutionary conserved core sets of self-propagating elements on ancient chromosomes of eukaryotes and metazoans. The same core TE set might be employed multiple times in evolution contributing to independent genome expansions. Beyond the highly conserved eukaryotic core set, our data also suggest the presence of several other metazoan core TE sets, including one that contains the remainder of hydra A-TEs.

Core TEs as slow-paced genomic expansion drivers

The distinction in abundance and repeat class composition of A-TE and non-A-TEs in animal and eukaryotic genomes point to two distinct regimes at which these elements can propagate. First, among the 82 investigated animal genomes, including those with the previously well described genome expansions, the variation of retroelement (LTR, LINE, SINE) contribution to genome sizes is most striking (Additional file 1: Fig. S10). Animals with “giant genomes” that have been sequenced so far show remarkable genome expansions by retroelements via copy-and-paste mechanisms [8, 9, 42, 84]. For example, axolotl (*Ambystoma mexicanum*) has a massive genome of 32 Gb, of which 65.6% consists of repetitive sequences, with LTRs accounting for 59% of the annotated repetitive elements [8]. The Australian lungfish (*Neoceratodus forsteri*) has a genome size of 37 Gb, with 90% being repetitive elements, and LINEs (mostly CR1 and L2 elements) dominating 25.7% of the genome [9]. Fixation of so many insertions, often contributing to substantial chromosome size increases, likely happens during periods of relaxed selection pressure, such as small population sizes or bottlenecks [85]. Conversely, our data demonstrate the presence of a much slower, but consistent chromosome expansion trend that is accomplished by the few A-TE families. Using comparison of the AEP and 105 strains, we show that these A-TEs can also propagate under normal sexual reproduction conditions. The functional role of A-TEs appears to be different depending on the stem cell lineage (Fig. 4, Additional file 1: Fig. S8). However, their effect may accumulate over macro-evolutionary timescales. In our investigation of chromosomal conformation capture, we observed that bins with higher A-TE coverage exhibited stronger chromatin interaction signals compared to those with lower A-TE coverage (Additional file 1: Fig. S11). Bins with higher A-TE coverage furthermore had significantly higher interactive ranges compared to other A-TE depleted bins (1.3 Mb vs 0.7 Mb, p value $< 2.2e - 16$, Wilcoxon rank sum test, Additional file 1: Fig. S11e). This may suggest that A-TEs could have much broader and subtle functions than individual gene expression control, being involved in overall chromatin organization, such as tethering elements through long-distance chromatin interactions [86] or other conformational states [87].

The role of A-TEs in the general chromosomal organization may also be supported by the observation that most, if not all, of the A-TE families are vertically transmitted

and have been retained independently in multiple animal genomes. This is supported by the total number of repeat copies present across many metazoan genomes that can be traced to these families (Fig. 5c and Additional file 1: Fig. S10). Additionally, repetitive elements that have been reported to be horizontally transferred in hydra (Hvmar1 and Hvmar2) [88] do not map to any of the A-TE families. A-TEs also show significantly greater sequence divergence compared to non-A-TEs (p value $< 2.2e - 16$, Wilcoxon rank sum test) and nested insertion patterns (Additional file 1: Fig. S12) [89], indicative of more ancient insertions. While we cannot rule out that some of the younger A-TE copies originated via horizontal transfer, our analyses indicate a general tendency of vertical transmission and retention of these families.

Discussion

How TEs have shaped and continue to affect genomes of animals remains an understudied question due to the large variance in genomic representation of these elements, and, more crucially, the lack of appropriate model systems to profile their activity and function. To address this question from both functional and evolutionary perspective, we utilized the cnidarian hydra with its distinct stem cell lineages that yield, under asexual reproduction, three independently evolving genomes within the same individual. With transgenic access to these genomes and cell type transcriptomics, we were able to reveal the putatively active TEs (A-TEs), the key TE components of these genomes and cell types, and to assess their expression and genomic insertion dynamics.

In this study, we provide direct evidence that the stem cells of hydra have diversified genomes due to specific sets of full-length and transcriptionally active TEs (A-TEs). Over a dozen of such TE families are actively and distinctively shaping genomic landscapes in interstitial, ectodermal, and endodermal stem cell lineages. Through insertion into sites in and around genes involved in various cellular functions, these insertions are likely to contribute to generating variation in expression of these genes. For example, the observation of TE insertions near genes involved in cell proliferation may suggest a race between TEs to make their host cell outcompete others. This mechanism may be of crucial selective advantage for an asexually reproducing animal such as hydra.

Leveraging the key phylogenetic position of hydra and using a broad eukaryotic species sampling, we also reveal that A-TEs are evolutionarily highly conserved either at the eukaryotic or metazoan levels, making them “core” evolutionarily preserved and active TEs of these genomes. This finding is striking, as so far deep homology of TEs has been mainly investigated at the superfamily level, but homology at family level has been largely inaccessible due to species sampling and the lack of high-quality genomic resources. Our findings suggest that only a few dozen TE families show ancient conservation and consistent genomic content across many deeply branching taxa. Other TE families either play a much smaller role in shaping genomes across larger phylogenetic scales or contribute to substantial genomic expansions in particular lineages. Our data suggest that genome expansions can be seen as a cumulative act of these distinct TE expansion regimes, generating a plethora of novel genomic material and driving chromosomal expansions. We also showed evidence that the A-TE aided expansion regime is more involved in long-range chromosomal contacts and thus may contribute to the overall chromosomal compaction.

At the metazoan level, beyond the core set dominated by the A-TEs identified in this study, the evolutionary comparison also hints at the existence of a few other core metazoan sets. We suggest that similar core sets from this widely conserved metazoan pool of TEs may exist and be employed in independent species-specific slow genome expansions. This suggests the possibility that animal chromosomes evolve through activation waves of particular core TE sets, while keeping the ancient chromosomal homologies stable. While the mechanistic insights into the evolutionary genome dynamics remains to be fully elucidated, the evolutionary conserved PIWI-piRNA pathway-mediated TE repression might be coupled with it [76]. Together, these findings shed new light on old questions on the vast TE diversity, modalities of metazoan genome evolution, and test how activations of core TE sets are contributing to divergent macro-evolutionary trajectories of animal genomes [90].

Conclusions

In this study, we investigated the activation and insertion dynamics of TEs across different stem cell lineages of hydra, a basally branching metazoan providing for a crucial comparative point to other animal genomes. We found that A-TEs, primarily composed of DNA elements, explain the minor genome size differences between different hydra strains and show distinct insertion profiles across the three hydra stem cell genomes. The majority of these TE families are evolutionarily conserved and are highly represented across most animal genomes. While these families are not the main contributors to major animal genome expansions, which are mainly driven by retroelements, we show evidence that insertions of A-TEs can impact gene expression, and that their accumulation over longer evolutionary timescales is correlated with stronger distal genomic contacts. This study provides evidence for a core set of TE families that has consistently accompanied animal chromosome evolution, opening up avenues for their further functional dissection.

Methods

Animal husbandry

Hydra strains were kept in individual containers filled with hydra medium at 16 °C. Hydroids were maintained under a continuous cycle of 14 h of light and 10 h of darkness. Hydroids received regular feedings of newly hatched brine shrimp. The hydra medium was changed daily. The Cnno1::GFP strain and the ecto-GFP/endo-RFP strain were from previous studies [72, 79].

Generation of the AEP × 105 strain

Males of the AEP Hym176B promoter::GFP line from Toshitaka Fujisawa were crossed with NE07a females. One of the progeny from this cross carried the transgene; this animal was used to establish a line. This line was heterozygous for the transgene. Females from this line were crossed with males from the Landing Site 6.1 line, which is an AEP transgenic line homozygous for the actin promoter::DsRed2 transgene and is six generations inbred. From this cross, lines were established that carried both transgenes (heterozygous for GFP and heterozygous for DsRed2). Males from these lines were crossed to females from the SS1 AEP line (an AEP F1 line that is “supersexy”). Subsequent studies

using progeny from these crosses showed that the transgenes were on the same chromosome and that in one of these progenies a crossover had occurred that linked the two transgenes on that chromosome. This line was 7/8 AEP and 1/8 NE07a, and it was heterozygous for the chromosome containing the linked transgenes. Males from this line were crossed with 105 females to generate the AEP $\times$ 105 hybrid strain. The AEP $\times$ 105 strain that was obtained did not carry the transgenes.

Nanopore sequencing for genome assembly

Extraction of high molecular weight DNA from a snap-frozen sample of hydra polyps of the AEP $\times$ 105 strain and Nanopore sequencing on an Oxford Nanopore Technologies PromethION were carried out by the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, which is supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Following the initial run, the flow cell was given a nuclease flush and reloaded for a second round of data collection. The two datasets were combined for subsequent use.

Omni-C sequencing

A snap-frozen sample of hydra polyps of the AEP $\times$ 105 strain was provided to Cantata Bio (Scotts Valley, CA) who produced and sequenced an Omni-C library.

Iso-Seq sequencing

RNA was extracted from hydra polyps of the AEP $\times$ 105 strain using the method of Chomczynski and Sacchi [91]. Following resuspension in DEPC-treated water, the RNA was precipitated with 2.5 M LiCl overnight at 4 °C. The RNA was pelleted, washed twice with 70% ethanol, resuspended in DEPC-treated water, and stored at – 80 °C. Synthesis of cDNA, preparation of the SMRTbell template, and sequencing (two SMRT cells) were done at the Genomics Research and Technology Hub (GRT Hub) at UC Irvine on a Pacific Bioscience Sequel II sequencer.

Generation of haplotype-resolved genome assembly of the AEP $\times$ 105 F1 hybrid

The heterozygosity of the AEP $\times$ 105 F1 hybrid genome was assessed from the Illumina short reads of the genome [1] using Jellyfish v2.3.1 [92], a k-mer counting software, with a k-mer size of 21. The k-mer spectrum was analyzed using GenomeScope [93]. The resulting k-mer profile displayed a single peak, suggesting that the AEP haplotype and the 105 haplotype were well diversified and resolving each haplotype genome is feasible (“pseudohomozygosity”). In the previous study, total 4420 contig sequences with the total size of 1.75 Gb were previously assembled from 23.6 Gb PacBio HiFi reads of the AEP $\times$ 105 genome [1]. The chromosome-level scaffolds were constructed with the combination of the 4420 contig sequences and the newly generated 26.2 Gb Omni-C (DNase I Hi-C) sequencing reads. The read quality of the Hi-C reads was evaluated using FastQC v0.12.1. The Hi-C reads were mapped to the contig sequences using Juicer v1.6 [94] with default parameters. Subsequently, Hi-C scaffolding was performed using the 3D-DNA pipeline v180922 [95] with parameters the default parameters followed by curation of resulting scaffolds using Juicebox Assembly Tools v1.11.08 [96]. This Hi-C scaffolding step produced 30 chromosomal-scale scaffolds which are corresponding to

both of 15 haplotype chromosome sets. These scaffolds comprise 97.1% (1.70 Gb) of the initial contigs. By comparing with previously reported genome assemblies of the AEP strain [60] and the 105 strain [1], we confirmed that 15 of the scaffolds correspond to the paternal AEP haplotype genome, while the remaining 15 correspond to the maternal 105 haplotype genome. The AEP haplotype genome sequences had 1523 gaps, while the 105 haplotype genome sequences had 2300 gaps. Subsequently, all these gaps were filled with the newly generated 48 Gb of raw Oxford Nanopore reads. We identified telomere reads from the previous HiFi reads and newly generated Nanopore reads for this study and extended the terminal sequences of chromosomes. Resulting scaffolds were polished with previously obtained 35.6 Gb Illumina short reads of the AEP × 105 F1 hybrid [1]. In the final genome assembly, each haplotype was gapless and featured telomere-subtelomere sequences at both ends of all 15 chromosomes. The AEP haplotype genome was 912 Mb in size, while the 105 haplotype genome was 834 Mb in size. Compared to previously reported data [1, 60], the AEP haplotype exhibited an increase of 10.9 Mb, while the 105 haplotype showed an increase of 16.2 Mb. Completeness of the genome assemblies examined using BUSCO v5.4.6 and the metazoa_odb10 datasets [97]. The BUSCO scores on genome sequences for both haplotype genomes were 91.7%, an improvement over the previously reported AEP genome assembly and 105 genome assembly scores of 91.5% and 91.4% [1, 60]. Based on these findings, the genome assembly appeared closer to a complete genome assembly. We have named this genome assembly HydraT2T. The sequences of the AEP haplotype genome and the 105 haplotype genome have been deposited in the NCBI Genome database under the accession numbers GCF_038396675.1 (HydraT2T_AEP) and GCF_037890685.1 (HydraT2T_105).

Generation of a custom repetitive element library

Since the majority of TEs are inactivated and degenerated due to accumulation of mutations such as base substitutions, insertions, and deletions, TE annotation in a pan-metazoan scale is challenging [98, 99]. At the first step of TE annotation, generating custom repeat libraries using RepeatModeler v2.0.5 [100] is a standard approach. A simple RepeatModeler run occasionally produces a library that contains lots of repetitive elements as “Unknown” elements. Analyzing the genome sequence with RepeatMasker v4.1.6 using this custom repeat library, which includes many “Unknown” sequences, inevitably results in a large number of repetitive elements being defined as “Unknown.” Since the process of RepeatModeler [100] generating consensus sequences for repetitive elements does not refer to the TE database such as Dfam [66], RepeatMasker using RepeatModeler custom library still detects the proper proportions of repetitive elements in the genome sequence. RepeatModeler classifies the consensus sequences of repetitive elements using its basic homology-based classification module called RepeatClassifier [100]. RepeatClassifier searches for homologous sequences within the Dfam sequences using the NCBI BLAST-based search (<https://github.com/Dfam-consortium/RepeatModeler>). If homologous sequences are not found in the database, RepeatMasker labels the remaining sequences as “Unknown” [100]. BLAST is a fast sequence search method that utilizes the dot matrix method. The hidden Markov model-based sequence search is

another sequence search approach and can be generally applied for homology searches for evolutionarily distant sequences [101, 102]. The hidden Markov model can even be used to detect RNA viral genes with highly divergent sequences [103, 104]. Therefore, for the consensus sequences generated by RepeatModeler, we performed annotation using the nhmmscan v3.4 [101], which uses a more sensitive hidden Markov model instead of BLAST, which RepeatClassifier uses. We searched for the best hit in the Dfam 3.8 database [66] for each consensus sequence of repetitive elements. For each consensus sequence of repetitive elements, the top hit from the nhmmscan search against Dfam was used as the TE annotation. Consensus sequences that did not have a hit identified in Dfam during the sequence search were labeled as “Unknown,” similar to how the RepeatClassifier operates. We have additionally validated our repeat library annotation using other external sources, including Repbase [67] and Gypsy Database [68].

TE detection in the genomes

With the custom repetitive element library, RepeatMasker [105] was performed on the genome sequences with the parameters of `-parallel 70 -gff -a -lib “repetitive element library file” -dir -xsmall`. For the benchmark analysis, we also generated another repetitive element library using RepeatModeler with default parameters. We collected chromosome-level genomic scaffolds of three unicellular holozoans, including *Creolimax fragrantissima*, *Capsaspora owczarzaki*, and *Salpingoeca rosetta* from the supplemental datasets in the previous study [2] (<https://doi.org/10.5061/dryad.dncjsxm47>). Then, we performed genome masking using RepeatMasker with two different repetitive element libraries: the custom repetitive library and the default repetitive library. The output files from RepeatMasker were processed using custom scripts, which computed various statistics, including TE coverage across the genome.

Analysis of nested insertion patterns for A-TEs

For the analysis of nested insertion patterns of A-TEs, the TinT tool [89] was used. The distribution of each A-TE was extracted from the RepeatMasker output and used as input for the TinT tool with default parameters to infer nested insertion relationships.

Genome annotation

To predict protein-coding genes in the genome assembly, we aligned Iso-Seq reads to the genome assemblies using the minimap2 v2.26-r1175 [106] with options `-ax splice -uf -C5`. We then performed genome annotation using BRAKER2 v 2.1.6 [107] on the repeat-masked genome assemblies masked by RepeatModeler [100] and RepeatMasker [105]. The UTR sequences were further added to the gene prediction with the parameters of `–addUTR = on –skipAllTraining`.

Syntenic analysis

Reciprocal blast best hits were used to define orthologs between species. The distribution of orthologs across species was plotted using the R package macrosyntR v0.3.3 with

default parameters [108]. Collinear syntenic blocks were identified using MCScanX with default parameters [109]. Collinear syntenic blocks were visualized using SynVisio [110].

Stem-cell lineage-specific whole-genome sequencing

One hundred hydra individuals with clonal propagation were carefully collected and transferred to a 2.0-ml tube containing dissociation medium [111]. The hydra were then gently dissociated by incubating them in Pronase solution (0.1% Pronase in hydra medium) for 90 min at 18 °C. Throughout the incubation phase, gentle agitation and pipetting were performed to facilitate the disassociation procedure. After the incubation, the Pronase activity was neutralized by adding an equal volume of BSA-containing hydra medium. The dissociated cells were subsequently washed three times with 1 ml dissociation medium. Finally, the dissociated cells were collected by centrifugation at 2000 rcf for 3 min at 4 °C and then resuspended in 1 ml dissociation medium. FACS were performed with BD FACSMelody™ Cell Sorter at the Max Perutz Labs BioOptics FACS Facility. The FACS-sorted cells were suspended in a lysis buffer (pH 8.0) containing 10 mM Tris, 100 mM EDTA, and 0.5% SDS to lysis the cell and solubilize genomic DNA. Equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) was added to the sample lysate, and the mixture was vortexed for 10 s to ensure proper emulsification. After mixing, the suspension was centrifuged at 12,000 rpm for 5 min at 4 °C to facilitate phase separation. The upper aqueous phase containing genomic DNA was carefully transferred to a new tube, avoiding the interphase. Equal volume of chloroform was added, and mixture was vortexed for 10 s. The suspension was centrifuged at 12,000 rpm for 5 min at 4 °C. The upper aqueous phase containing genomic DNA was carefully transferred to a new tube, and the same step was repeated once. The final aqueous phase was treated with equal volume of 2.5 volumes of room temperature isopropanol and 1/10 volume of 3 M sodium acetate (pH 5.5). The sample was mixed well. After incubation, the sample was centrifuged at 6000 rcf for 5 min at 4 °C. The supernatant was discarded, and the pellet was washed three times with 70% ethanol, air-dried, and finally resuspended in 100 µl TE buffer. The concentration and purity of the extracted genomic DNA were determined using a Nanodrop spectrophotometer. Library preparation and nanopore sequencing were performed at the Next Generation Sequencing core facility at the Vienna BioCenter.

Nanopore read mapping to the genome sequence and variant calling for the stem-cell lineage whole-genome sequencings

Qualities of the Nanopore reads were evaluated using NanoPlot [112]. Nanopore reads were aligned to the hydra reference genome sequence [60] using minimap2 v2.26-r1175 with options of -ax map-ont -secondary=no [106]. Variant calling for each individual sample was performed using Sniffles v2.02 [113] with the -minsupport 2 option which produces a SNF file for each sample. Individual variant calls were merged using Sniffles v2.02 with options of -combine-low-confidence 0.01 -combine-low-confidence-abs 1 -combine-null-min-coverage 3 which produce a combined VCF file. The genomic insertions where all samples have genotypes were kept using bcftool v1.19 [114]. The insertions were further subset by bcftools v1.19 based on the combination of genotypes

for each sample, for the downstream analyses. For insertion annotations, sequences of insertions longer than 1 kb were extracted and searched against the RepeatModeler repeat library using NCBI BLASTN [115]. The TE annotation of an insertion was determined based on the consensus sequence of the custom repeat library that formed an alignment of 500 bp or more with the TE sequence.

Functional gene analysis

For functional gene analysis, protein sequences were retrieved from the genome sequence and the genome annotations using GffRead v0.12.7 [116]. The longest protein sequences were analyzed to identify homologous sequences within the NCBI-nr database, employing the NCBI BLASTN program with an *e*-value threshold of $1e-05$. The longest protein sequence for each gene was functionally annotated using eggNOG-Mapper v2.1.9 [117]. Genes proximal to TE insertion sites were identified using BEDTools v2.30.0 [118] with a subcommand of closest. We conducted Fisher's exact test for the overrepresentation of GO terms within the group of genes proximal to TE insertion sites, utilizing the R library topGO v 2.54.0 [119]. Results of the statistical tests were plotted using the R library ggplot2 v 3.5.0 [120].

RNA-seq analysis

To determine whether each Iso-Seq read is derived from TEs, each read sequence from Iso-Seq was searched against the repeat library generated by RepeatModeler [100] using NCBI BLASTN [115] with default parameters. Iso-Seq reads that formed alignments of over 1 kb with the sequences in the repeat library were called as TE-derived sequences. Furthermore, Iso-Seq reads defined as TE-derived were mapped to the genome sequence using Minimap2 v2.26-r1175 with parameters of `-ax map-hifi --secondary=no`, and the mapped regions were identified. Additionally, to reveal the transcription of TEs at a single-cell resolution, sequencing reads from a publicly available scRNA-seq dataset of hydra [54] were downloaded. From this dataset, only TE sequences belonging to the TE sub-families (RepeatModeler consensus sequences) confirmed to be expressed by Iso-Seq and that were over 1 kb in length were used. With these TE sequences, the expression levels of TEs in each cell were quantified using Alevin v1.2.1 [121]. The resulting expression matrix was loaded using the CreateSeuratObject function of the R package Seurat v5 [122]. Cell type annotation was based on the previous reports [54, 60]. TEs showing cell type-specific expression were identified using the FindAllMarkers function of Seurat. Additionally, RNA-seq datasets for FACS-sorted cell populations [53] were mapped to the genome sequence using HISAT2 v 2.2.1 [123]. The read counts of each TE were evaluated with featureCounts [124], and transcripts per million were calculated for each TE. The expression of TEs showing cell type-specific expression obtained from scRNA-seq was checked. Heatmaps were plotted using the R package ComplexHeatmap v 2.18.0 [125].

Identification of open chromatin regions

To identify open chromatin regions in the hydra genome, publicly available ATAC-seq reads (NCBI accession number: SRR18362011) from whole adult polypos were obtained

from the NCBI SRA database. These reads were aligned to the hydra genome assembly using BWA-MEM [126]. Peak calling from the ATAC-seq read alignment was performed using MACS3 v 3.0.2 [127]. Overlapping TEs and ATAC-seq reads were analyzed using BEDTools v2.30.0 [118] with a subcommand of intersect.

Analysis of chromatin organization

Omni-C reads were aligned to the hydra genome assembly, and chromatin contact intensities were quantified using the Juicer/3D-DNA pipeline [96]. HiC contact intensities were extracted with hic-straw v 1.3.1 [96] and then converted into a BED file. The overlap between HiC bins and TE loci was identified using BEDTools v2.30.0 [118] with a subcommand of intersect. Statistical tests and visualization of results were performed using R v4.3.2.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-025-03653-z>.

Additional file 1: Figures S1–S12, Supplementary Notes 1–2.

Additional file 2: Tables S1–S15.

Acknowledgements

The authors would like to express their gratitude to Yoshihiro Omori from the Hiroshima University, Japan, for his advice on genome extraction and to Juan Daniel Montenegro Cabrera of the University of Vienna, Austria, for his comments regarding repeat annotation of the Hydra genome. Our appreciation also extends to Nina Znidaric, Angela Caballero Alfonso, and Eduard Renfer of the University of Vienna for their dedication to animal care and technical assistance, as well as to Chiemi Nishimiya-Fujisawa from Heidelberg University, Germany, for her valuable advice on Hydra strains. We are thankful to Bert Hobmayer of the University of Innsbruck, Austria, for his insightful comments on this project. Whole genome sequencing was expertly performed by the Next Generation Sequencing core facility at the Vienna BioCenter. Special thanks to Kitti Dora Csalyi and Thomas Sauer at the Max Perutz Labs BioOptics FACS Facility for their valuable advice on the FACS experiments. Most of the computations were carried out on the Life Science Compute Cluster (LiSC) of the University of Vienna. R.E.S. thanks Dr. Ira Blitz (UC Irvine) for his assistance with RNA isolation for Iso-Seq sequencing. We thank all members of the Simakov lab, especially Darrin Schultz for valuable comments on the manuscript.

Peer review information

Jürgen Schmitz and Tim Sands were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions

KKN, TK, ET, TH and OS designed the research. RS and TH provided materials. KKN, TK, and OS generated the genome assemblies. KKN, TK, OS, TKY, DRT, FF, DM, RS, ET, and TH performed the analyses. KKN, TK, and OS wrote and revised the manuscript with comments from the other authors. All authors read and approved the final manuscript.

Funding

This work was supported by the Austrian Science Fund FWF (I4353) to OS and EMT, D-A-CH (Ho1088-10-1) to TWH, Takeda Science Foundation to TK, the Mochida Memorial Foundation for Medical and Pharmaceutical Research to TK, the International Medical Research Foundation to TK, and the Yamada Science Foundation to TK. OS was supported by an ERC-H2020-EURIP grant, grant number 945026. This work utilized resources of the UCI Genomics Research and Technology Hub (GRT Hub) parts of which are supported by NIH grants to the Comprehensive Cancer Center (P30CA-062203) and the UCI Skin Biology Resource Based Center (P30AR075047) at the University of California, Irvine, as well as to the GRT Hub for instrumentation (1S10OD010794-01 and 1S10OD021718-01).

Data availability

The sequencing reads are available in the NCBI database under BioProject accession number PRJNA1085048. The genome assemblies for the AEP and 105 genomes are available in the NCBI database under BioProject accession numbers PRJNA1085304 and PRJNA1085315, respectively. The scripts for this manuscript are available at the Zenodo repository (<https://doi.org/10.5281/zenodo.15600058>) [128]. The RepeatMasker output files generated in this study are publicly available on the Figshare repository (<https://doi.org/10.6084/m9.figshare.28775144.v2>) [129]. The GenBank accession numbers of the publicly available ITS1-5.8S-ITS2 rDNA sequences used in this study are listed in Table S3 of Additional file 2. The NCBI genome accession number corresponding to the publicly available genome sequences analyzed in this study is listed in Table S10 of Additional file 2.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 January 2025 Accepted: 13 June 2025

Published online: 01 July 2025

References

- Simakov O, et al. Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci Adv*. 2022;8:eabi5884.
- Schultz DT, et al. Ancient gene linkages support ctenophores as sister to other animals. *Nature*. 2023;618:110–7.
- Simakov O, et al. Insights into bilaterian evolution from three spiralian genomes. *Nature*. 2013;493:526–31.
- Zimmermann B, Robert NSM, Technau U, Simakov O. Ancient animal genome architecture reflects cell type identities. *Nat Ecol Evol*. 2019;3:1289–93.
- Robert NSM, et al. Emergence of distinct syntenic density regimes is associated with early metazoan genomic transitions. *BMC Genomics*. 2022;23:143.
- Wilch ES, Morton CC. Historical and clinical perspectives on chromosomal translocations. *Adv Exp Med Biol*. 2018;1044:1–14.
- Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet*. 2007;41:331–68.
- Nowoshilow S, et al. The axolotl genome and the evolution of key tissue formation regulators. *Nature*. 2018;554:50–5.
- Meyer A, et al. Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature*. 2021;590:284–9.
- Sun C, Mueller RL. Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. *Genome Biol Evol*. 2014;6:1818–29.
- Kapusta A, Suh A & Feschotte C. Dynamics of genome size evolution in birds and mammals. *Proc Natl Acad Sci*. 2017;114:E1460–E1469.
- Tenaillon M, Hollister JD, Gaut BS. A triptych of the evolution of plant transposable elements. *Trends Plant Sci*. 2010;15:471–8.
- DeBarry JD, Kissinger JC. Jumbled genomes: missing apicomplexan synteny. *Mol Biol Evol*. 2011;28:2855–71.
- Kissinger JC, DeBarry J. Genome cartography: charting the apicomplexan genome. *Trends Parasitol*. 2011;27:345–54.
- McVean G. What drives recombination hotspots to repeat DNA in humans? *Philos. Trans R Soc B Biol Sci*. 2010;365:1213–8.
- O'Neill RJW, O'Neill MJ, Graves JAM. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature*. 1998;393:68–72.
- Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet*. 2017;18:71–86.
- Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 2008;9:397–405.
- Emera D, Wagner GP. Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. *Proc Natl Acad Sci*. 2012;109:11246–51.
- Papolu PK, et al. Retrotransposons: how the continuous evolutionary front shapes plant genomes for response to heat stress. *Front Plant Sci*. 2022;13:1064847.
- Gianfrancesco O, Bubba VJ, Quinn JP. SVA retrotransposons as potential modulators of neuropeptide gene expression. *Neuropeptides*. 2017;64:3–7.
- Du AY, Chobirko JD, Zhuo X, Feschotte C, Wang T. Regulatory transposable elements in the encyclopedia of DNA elements. *Nat Commun*. 2024;15:7594.
- Fueyo R, Judd J, Feschotte C, Wysocka J. Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol*. 2022;23:481–97.
- Wells JN, et al. Transposable elements drive the evolution of metazoan zinc finger genes. *Genome Res*. 2023;33:1325–39.
- Eichler EE, et al. Complex β -satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. *Genome Res*. 1998;8:791–808.
- Kapusta A, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet*. 2013;9: e1003470.
- Chang N-C, et al. Gag proteins encoded by endogenous retroviruses are required for zebrafish development. *bioRxiv*. 2024.
- Cosby RL, et al. Recurrent evolution of vertebrate transcription factors by transposase capture. *Science*. 2021;371:80.
- Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. *Annu Rev Genet*. 2020;54:539–61.
- Lu JY, et al. Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res*. 2021;31:613–30.

31. Li W, et al. Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nat Neurosci*. 2013;16:529–31.
32. Petersen HO, et al. A comprehensive transcriptomic and proteomic analysis of hydra head regeneration. *Mol Biol Evol*. 2015;32:1928–47.
33. Albertin CB, et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature*. 2015;524:220–4.
34. Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Mol Biol Evol*. 2014;31:1188–99.
35. Putnam NH, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*. 2007;317:86–94.
36. Chapman JA, et al. The dynamic genome of *Hydra*. *Nature*. 2010;464:592–6.
37. Bourque G, et al. Ten things you should know about transposable elements. *Genome Biol*. 2018;19:199.
38. Naville M, et al. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr Biol*. 2019;29:1161–1168.e6.
39. Blommaert J, Riss S, Hecox-Lea B, Mark Welch DB, Stelzer CP. Small, but surprisingly repetitive genomes: transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species complex. *BMC Genomics*. 2019;20:466.
40. Wong WY, et al. Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus *Hydra*. *Proc Natl Acad Sci U S A*. 2019;116:22915–7.
41. Gozashti L, Feschotte C & Hoekstra HE. Transposable element interactions shape the ecology of the deer mouse genome. *Mol Biol Evol*. 2023;40:msad069.
42. Schartl M, et al. The genomes of all lungfish inform on genome expansion and tetrapod evolution. *Nature*. 2024;634:96–103.
43. Erwin DH. Early metazoan life: divergence, environment and ecology. *Philos Trans R Soc B Biol Sci*. 2015;370:20150036.
44. Parry LA, et al. Ichneological evidence for meiofaunal bilaterians from the terminal Ediacaran and earliest Cambrian of Brazil. *Nat Ecol Evol*. 2017;1:1455–64.
45. Trembley A. *Mémoires, pour servir à l'histoire d'un genre de polypes d'eau douce, à bras en forme de cornes*. (Jean and Herman Verbeek, 1744).
46. Browne EN. The production of new hydranths in *Hydra* by the insertion of small grafts. *J Exp Zool*. 1909;7:1–23.
47. Lenhoff HM. Ethel Browne, Hans Spemann, and the discovery of the organizer phenomenon. *Biol Bull*. 1991;181:72–80.
48. Holstein TW. The significance of Ethel Browne's research on *Hydra* for the organizer concept. *Cells Dev*. 2024;178:203907.
49. Holstein TW. The role of cnidarian developmental biology in unraveling axis formation and Wnt signaling. *Dev Biol*. 2022;487:74–98.
50. Martínez DE. Mortality patterns suggest lack of senescence in *hydra*. *Exp Gerontol*. 1998;33:217–25.
51. Schaible R, et al. Constant mortality and fertility over age in *Hydra*. *Proc Natl Acad Sci*. 2015;112:15701–6.
52. Vogg MC, Galliot B & Tsiairis CD. Model systems for regeneration: *Hydra*. *Development*. 2019;146:dev177212.
53. Nishimiya-Fujisawa, C. et al. An ancient split of germline and somatic stem cell lineages in *Hydra*. *bioRxiv*. 2023.
54. Siebert, S. et al. Stem cell differentiation trajectories in *Hydra* resolved at single-cell resolution. *Science*. 2019;365:eaav9314.
55. Holzem M, Boutros M, Holstein TW. The origin and evolution of Wnt signalling. *Nat Rev Genet*. 2024;25:500–12.
56. Hobmayer B, et al. WNT signalling molecules act in axis formation in the diploblastic metazoan *Hydra*. *Nature*. 2000;407:186–9.
57. Buss LW. The evolution of individuality. Princeton University Press; 1988. ISBN 978-0691084688.
58. Martínez DE, et al. Phylogeny and biogeography of *Hydra* (Cnidaria: Hydridae) using mitochondrial and nuclear DNA sequences. *Mol Phylogenet Evol*. 2010;57:403–10.
59. Hamada M, Satoh N & Khalturin K. A reference genome from the symbiotic hydrozoan, *Hydra viridissima*. *G3 (Bethesda)*. 2020;10:3883–3895.
60. Cazet JF, et al. A chromosome-scale epigenetic map of the *Hydra* genome reveals conserved regulators of cell state. *Genome Res*. 2023;33:283–298.
61. Miller MA, Technau U, Smith KM, Steele RE. Oocyte development in *Hydra* involves selection from competent precursor cells. *Dev Biol*. 2000;224:326–38.
62. Martin VJ, Littlefield CL, Archer WE, Bode HR. Embryogenesis in *hydra*. *Biol Bull*. 1997;192:345–63.
63. Juliano CE, Lin H & Steele RE. Generation of transgenic *Hydra* by embryo microinjection. *J Vis Exp*. 2014;e51888.
64. Sugiyama T, Fujisawa T. Genetic analysis of developmental mechanisms in *Hydra*. II. Isolation and characterization of an interstitial cell-deficient strain. *J Cell Sci*. 1978;29:35–52.
65. Storer JM, Hubley R, Rosen J & Smit AFA. Curation guidelines for de novo generated transposable element families. *Curr Protoc*. 2021;1:e154.
66. 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;12:2.
67. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015;6:11.
68. Llorens C, et al. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res*. 2011;39:D70–4.
69. Macías-Muñoz A, Mesrop LY, Liang HY & Mortazavi A. Single cell analysis reveals the molecular signaling and cellular composition of the regenerating *Hydra* head. *bioRxiv*. 2022.
70. Stover NA, Steele RE. Trans-spliced leader addition to mRNAs in a cnidarian. *Proc Natl Acad Sci U S A*. 2001;98:5693–8.
71. Stover NA, Kaye MS, Cavalcanti ARO. Spliced leader trans-splicing. *Curr Biol*. 2006;16:R8–9.

72. Nishimiya-Fujisawa C, Kobayashi S. Germline stem cells and sex determination in Hydra. *Int J Dev Biol*. 2012;56:499–508.
73. Campbell RD, David CN. Cell cycle kinetics and development of Hydra attenuata: II. Interstitial cells *J Cell Sci*. 1974;16:349–58.
74. David CN, Campbell RD. Cell cycle kinetics and development of Hydra attenuata: I. Epithelial cells *J Cell Sci*. 1972;11:557–68.
75. Juliano CE, et al. PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. *Proc Natl Acad Sci U S A*. 2014;111:337–42.
76. Teefy BB, Siebert S, Cazet JF, Lin H, Juliano CE. PIWI-piRNA pathway-mediated transposable element repression in Hydra somatic stem cells. *RNA*. 2020;26:550–63.
77. Gagliardi PA, di Blasio L, Primo L. PDK1: a signaling hub for cell migration and tumor invasion. *Biochim Biophys Acta*. 2015;1856:178–88.
78. Raimondi C, Falasca M. Targeting PDK1 in cancer. *Curr Med Chem*. 2011;18:2763–9.
79. Glauber KM, et al. A small molecule screen identifies a novel compound that induces a homeotic transformation in Hydra. *Development*. 2015;142:2081.
80. Keeling PJ, Burki F. Progress towards the tree of eukaryotes. *Curr Biol*. 2019;29:R808–17.
81. Roger AJ, Simpson AGB. Evolution: revisiting the root of the eukaryote tree. *Curr Biol*. 2009;19:R165–7.
82. Howe K, et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*. 2013;496:498–503.
83. Marino A, Kizenko A, Wong WY, Ghiselli F, Simakov O. Repeat age decomposition informs an ancient set of repeats associated with coleoid cephalopod divergence. *Front Genet*. 2022;13: 793734.
84. Schloissnig S, et al. The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. *Proc Natl Acad Sci*. 2021;118:e2017176118.
85. Lynch M, Conery JS. The origins of genome complexity. *Science*. 2003;302:1401–4.
86. Batut PJ, et al. Genome organization controls transcriptional dynamics during development. *Science*. 2022;375:566–70.
87. Campos-Sánchez R, Kapusta A, Feschotte C, Chiaromonte F, Makova KD. Genomic landscape of human, bat, and ex vivo DNA transposon integrations. *Mol Biol Evol*. 2014;31:1816–32.
88. Grace CA, Carr M. The evolutionary history of mariner elements in stalk-eyed flies reveals the horizontal transfer of transposons from insects into the genome of the cnidarian Hydra vulgaris. *PLoS ONE*. 2020;15: e0235984.
89. Churakov G, et al. A novel web-based TinT application and the chronology of the primate Alu retroposon activity. *BMC Evol Biol*. 2010;10:376.
90. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature*. 1980;284:601–3.
91. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem*. 1987;162:156–9.
92. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27:764–70.
93. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432.
94. Durand NC, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016;3:95–8.
95. Dudchenko O, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356:92–5.
96. Durand NC, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 2016;3:99–101.
97. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2.
98. Goubert C, et al. A beginner's guide to manual curation of transposable elements. *Mob DNA*. 2022;13:7.
99. Matsuhashi W, Planet E, Trono D. Ancestral genome reconstruction enhances transposable element annotation by identifying degenerate integrants. *Cell genomics*. 2024;4: 100497.
100. Flynn JM, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020;117:9451–7.
101. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. *Bioinformatics*. 2013;29:2487–9.
102. Skewes-Cox P, Sharpton TJ, Pollard KS, DeRisi JL. Profile hidden Markov models for the detection of viruses within metagenomic sequence data. *PLoS ONE*. 2014;9: e105067.
103. Nakagawa S, et al. Current trends in RNA virus detection through metatranscriptome sequencing data. *FEBS Open Bio*. 2023;13:992–1000.
104. Wolf YI, et al. Origins and evolution of the global RNA virome. *MBio*. 2018;9:e02329–18.
105. Tarailo-Graovac M, & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinforma*. 2009;Chapter 4:4.10.1–4.10.14.
106. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–100.
107. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics Bioinforma*. 2021;3:lqaa108.
108. Hilali S El & Copley RR. macrosyntR: drawing automatically ordered Oxford Grids from standard genomic files in R. *bioRxiv*. 2023.
109. Wang Y, et al. MCSscanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40: e49.
110. Bandi V, Gutwin C. Interactive exploration of genomic conservation. *Proc Graph Interface*. 2020;2020:74–83.
111. Gierer A, et al. Regeneration of hydra from reaggregated cells. *Nat New Biol*. 1972;239:98–101.
112. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*. 2018;34:2666–9.

113. Sedlazeck FJ, et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*. 2018;15:461–8.
114. Danecek P, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10.
115. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
116. Pertea G & Pertea M. GFF Utilities: GffRead and GffCompare. *F1000Research* 2020;9.
117. Huerta-Cepas J, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47:D309–14.
118. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
119. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006;22:1600–7.
120. Wickham, H. *ggplot2*. (Springer New York, 2009). <https://doi.org/10.1007/978-0-387-98141-3>.
121. Srivastava A, Malik L, Smith T, Sudbery I, Patro R. Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. *Genome Biol*. 2019;20:65.
122. Hao Y, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. 2024;42:293–304.
123. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12:357–60.
124. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30.
125. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847–9.
126. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
127. Zhang Y, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137.
128. Kon T. Hydra transposable element analyses. 2025. Zenodo. <https://doi.org/10.5281/zenodo.15600058>.
129. Kon T. Custom repetitive element libraries for hydra genomes. 2025. Figshare. <https://doi.org/10.6084/m9.figshare.28775144.v2>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

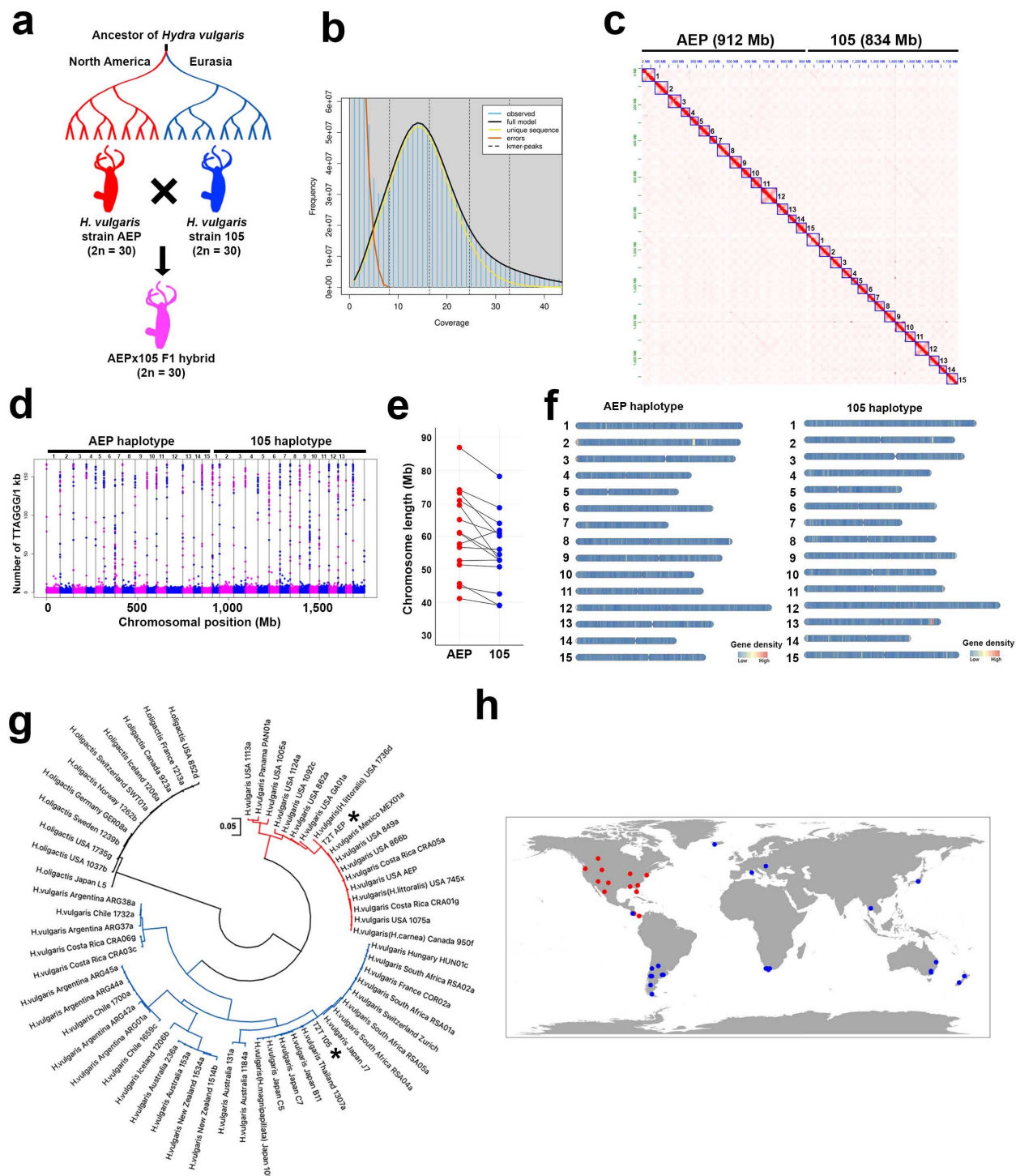


Fig. S1. The 30 chromosomes of the AEPx105 F1 hybrid

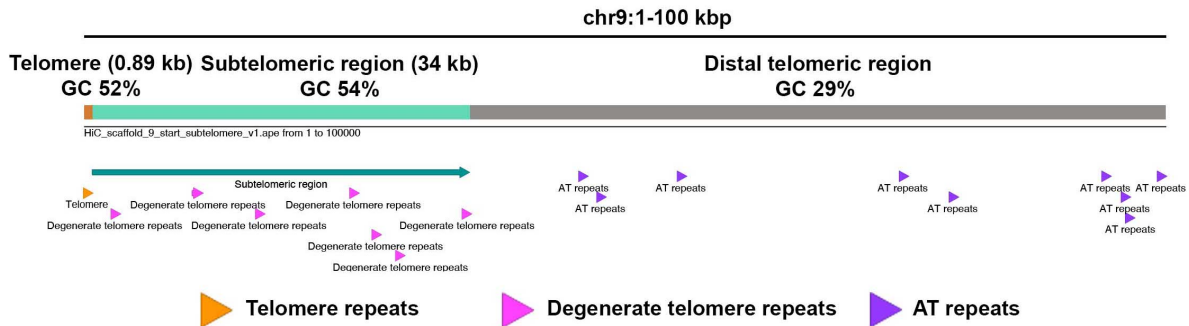
(a) Generation of AEPx105 F1 hybrid by crossing the AEP strain and the 105 strain. (b) The k-mer frequency distribution of the AEPx105 F1 hybrid genome. The k-mer ($k = 21$) occurrence was analyzed in the raw short-reads of the AEPx105 F1 hybrid genome¹. (c) Hi-C contact map of the 30 chromosomes of the AEPx105 F1 hybrid. In each chromosome, interaction signals at both ends of the chromosome are observed, suggesting a Rabl-like configuration, an arrangement of

interphase chromosomes where centromeres and telomeres are located at opposing poles within the nucleus. **(d)** Distribution of telomeric repeats. The x-axis represents genomic positions and the y-axis shows the number of occurrences of telomeric repeats (TTAGGG) per kilobase pair. The plots of each chromosome are colored alternately to distinguish between the plots of the terminal regions of two different chromosomes. **(e)** Chromosome lengths of each haplotype genome. Plots between orthologous chromosomes are linked with black lines. **(f)** Chromosomes are colored according to gene density. The positions of the centromeres are indicated by notches. It is evident that each haplotype genome consists of metacentric or submetacentric chromosome. **(g)** Molecular phylogenetic positions of the AEP haplotype and the 105 haplotype. Maximum-likelihood tree based on sequences of the internal transcribed spacer (ITS) 1, the 5.8S ribosomal RNA gene, and ITS2 from the AEP haplotype, the 105 haplotype, and sequences from a public database. Samples of *Hydra oligactis* are colored black. Two major clades of *Hydra vulgaris* are colored red and blue. The ITS1-5.8S-ITS2 sequences of the AEP haplotype and 105 haplotype are indicated by asterisks. **(h)** Geographic distribution of the samples whose ITS1-5.8S-ITS2 sequences are publicly-available. Each dot represents individual sample, colored in accordance with the Maximum-likelihood tree.

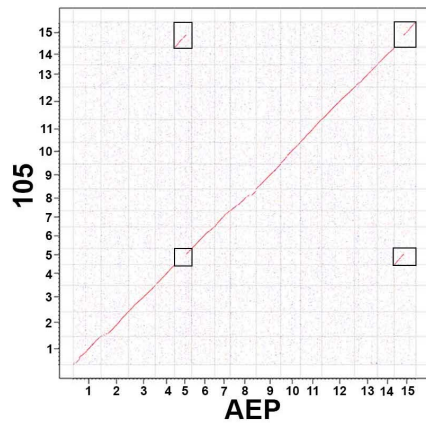
a



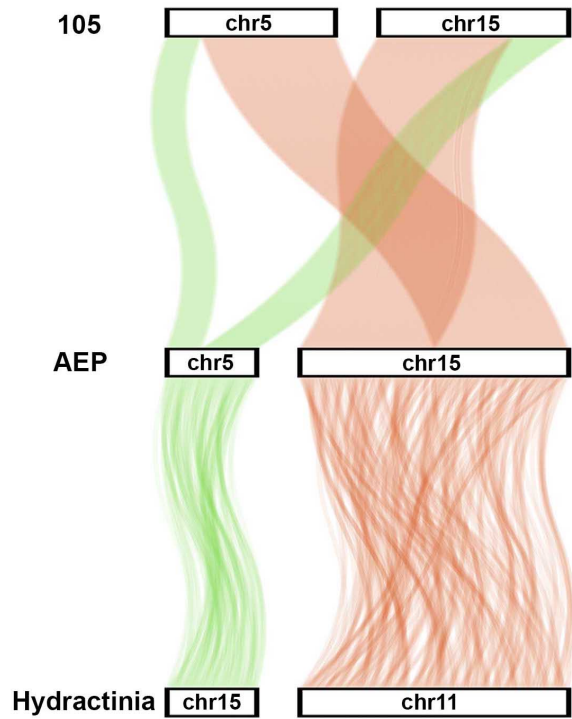
b



c



e



d

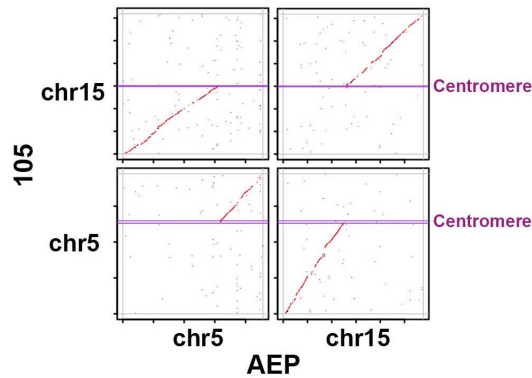


Fig. S2. The structure of hydra telomeric, subtelomeric and centromeric regions

(a) Telomere and subtelomere sequences in the 1-4,361 bp region of the chromosome 9. **(b)** Motif map of the telomere, subtelomere, and distal telomeric regions in chromosome 9. **(c)** Genome alignment between the AEP chromosomes and the 105 chromosomes. Alignments of the chromosomes involved in the translocation is highlighted with rectangles. **(d)** Genome alignment between chromosome 5 and chromosome 15 from the AEP haplotype genome and those from the 105 haplotype. The start and end of the centromeric regions are indicated by horizontal lines. **(e)** Highlight of the Robertsonian translocation. Each curved line indicates positions of orthologs.

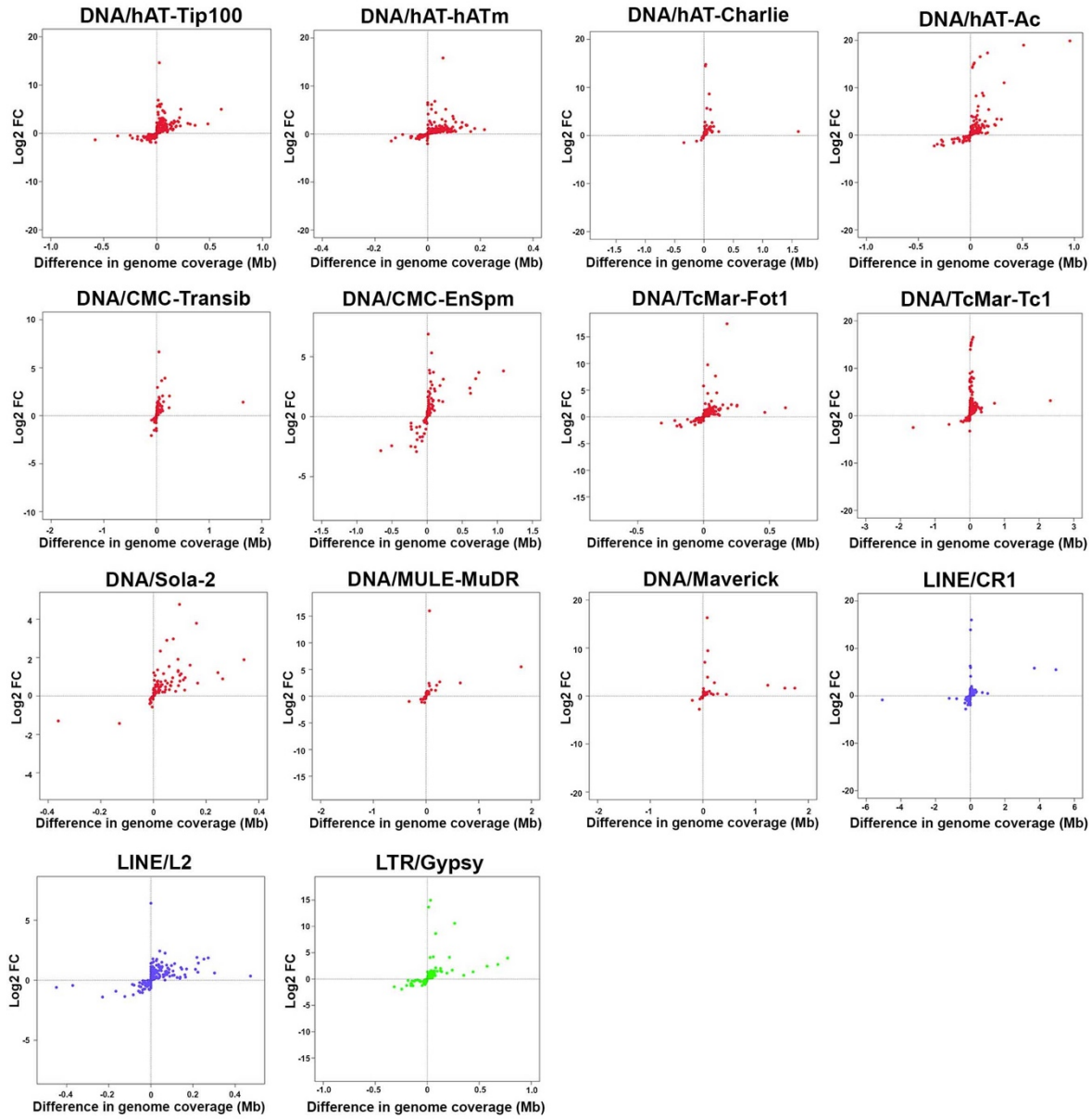


Fig. S3. Divergence of the genome coverage of subfamilies of A-TEs between the AEP and 105 genomes

The x-axis represents the difference in genome coverage of each TE subfamily between the AEP and 105 genomes (in Mb), while the y-axis displays the log₂-transformed fold change (Log₂ FC) in genome coverage of each TE subfamily between the two genomes. Each dot represents each subfamily.



Fig. S4. Divergence of TE subfamilies between the AEP and 105 genomes

(a) Collinear syntenic blocks between the 105 chromosomes and the AEP chromosomes. Collinear syntenic blocks are shown as ribbon diagrams. Syntenic blocks with same chromosomal orientations are colored in blue. Syntenic blocks with reverse orientations are colored in red. The

length of the chromosomes is depicted as proportion to the assembly size of each chromosome. **(b)** Distributions of TE subfamilies across chromosomes. The leftmost panel shows the distribution of collinear syntenic regions in black. The second panel from the left represents the density plot for the 14 A-TEs. The remaining panels provide examples of the distributions of subfamilies of A-TEs. The red lines indicate genomic regions where collinear syntenic regions were not detected.

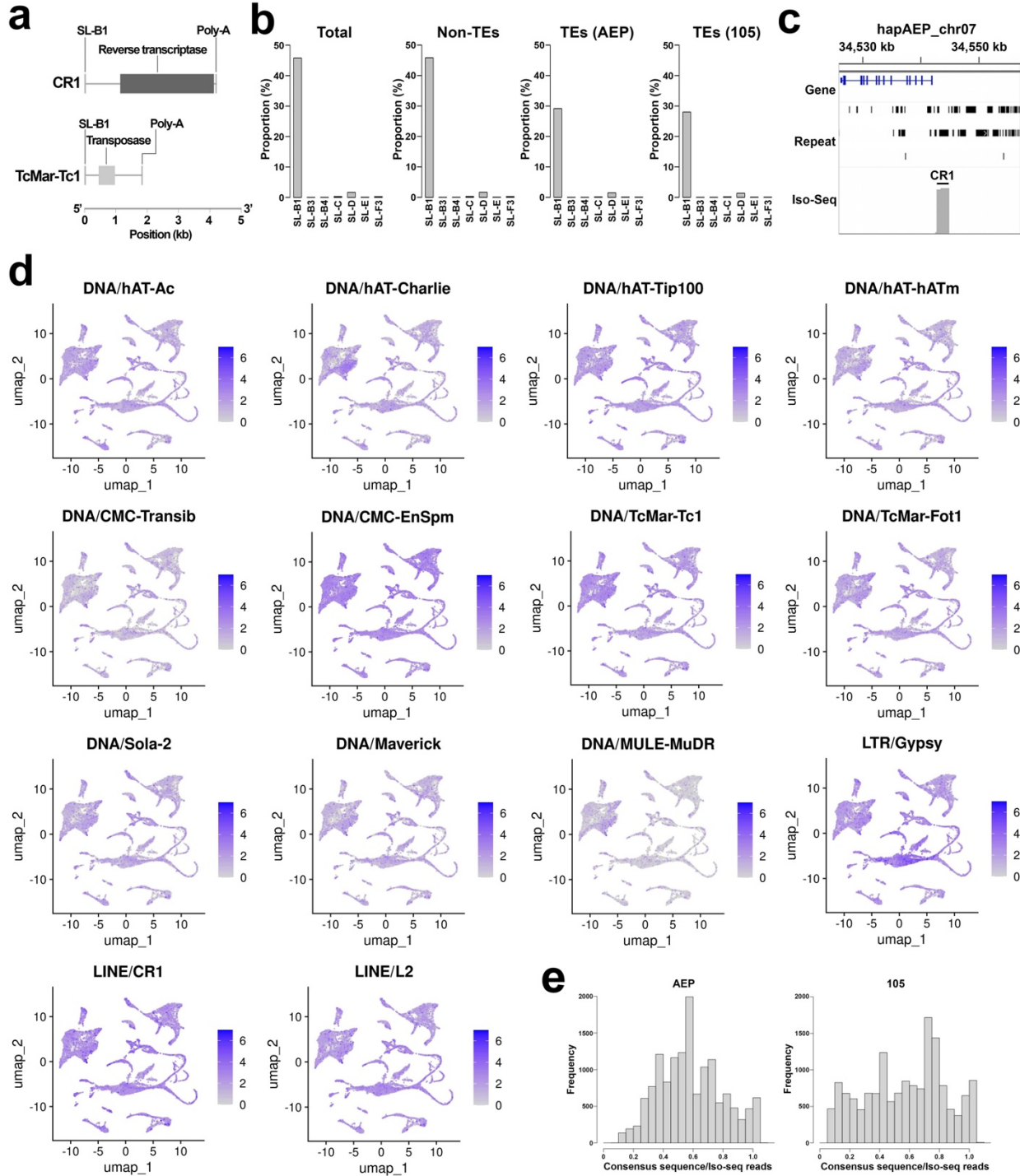


Fig. S5. Expression profiles of TE families

(a) Full length transcripts of the CR1 element and TcMar-Tc1 element with trans-spliced leader sequences (SL-B1²) attached at the 5' end. **(b)** Proportions of transcripts with trans-spliced leader sequences revealed by Iso-Seq. The x-axes represent types of trans-spliced leader sequences². The y-axes represent the proportions of transcripts with trans-spliced leader sequences. **(c)** Representative locus with CR1 expression measured by Iso-Seq. **(d)** Expression profiles of the A-TEs at the family level. The color scale indicating gene expression levels is the same in all panels,

and the variations in color intensity are comparable across different panels. **(e)** Coverage of TE subfamily sequences in TE Iso-seq reads.

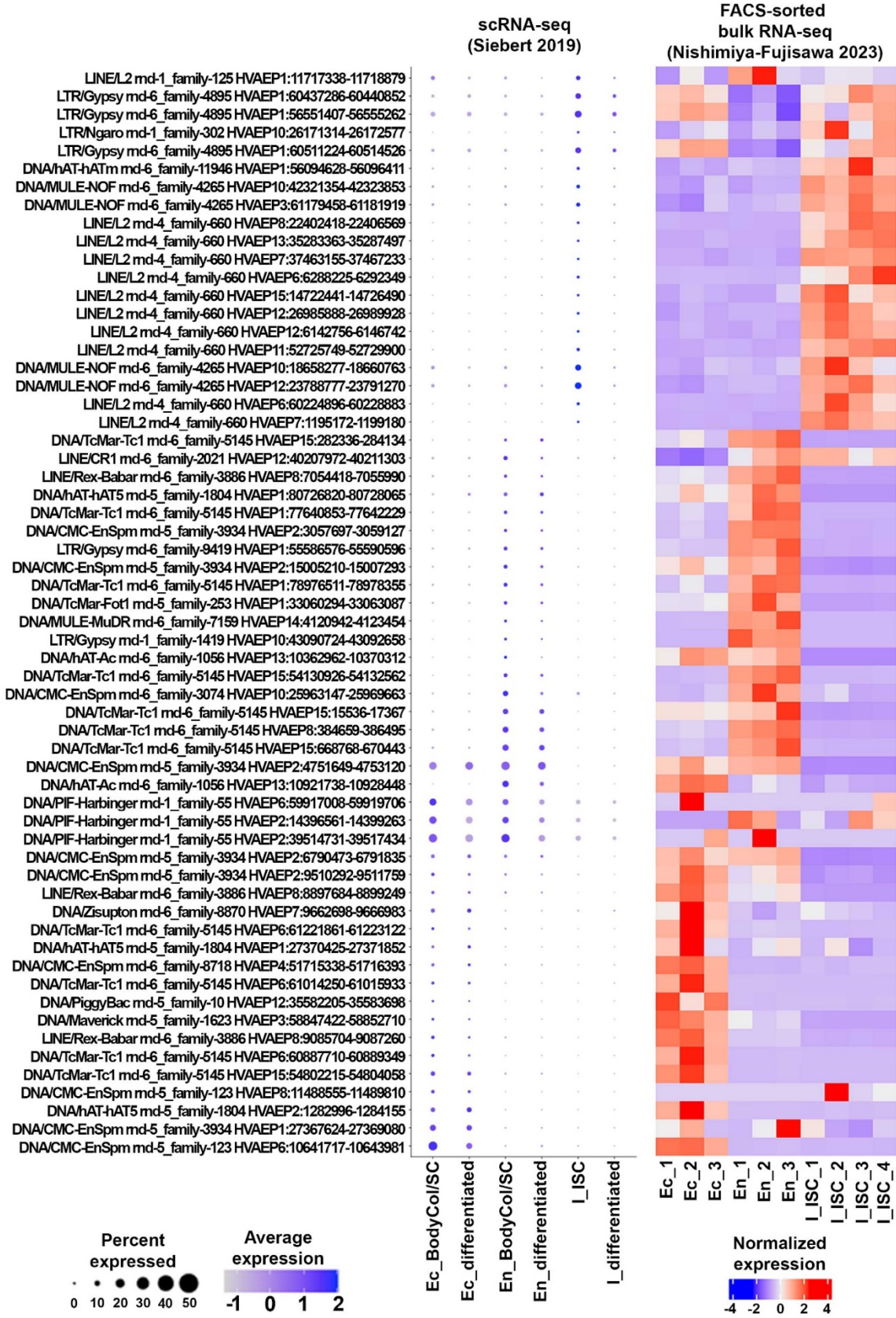


Fig. S6. Expression profiles of TEs at locus level

Expression profiles of TEs at locus level were investigated using two different datasets^{3,4}. In each group of ectodermal stem cells (Ec_BodyCol/SC), endodermal stem cells (En_BodyCol/SC), and i-cells (I_ISC), top20 TEs that were significantly more expressed are shown. Average expression levels (color intensity) and the proportion of cells expressing them (circle size) are shown. Additionally, as a validation, the expression levels of these TEs in a FACS-sorted bulk RNA-seq dataset are displayed on a heatmap to the right.

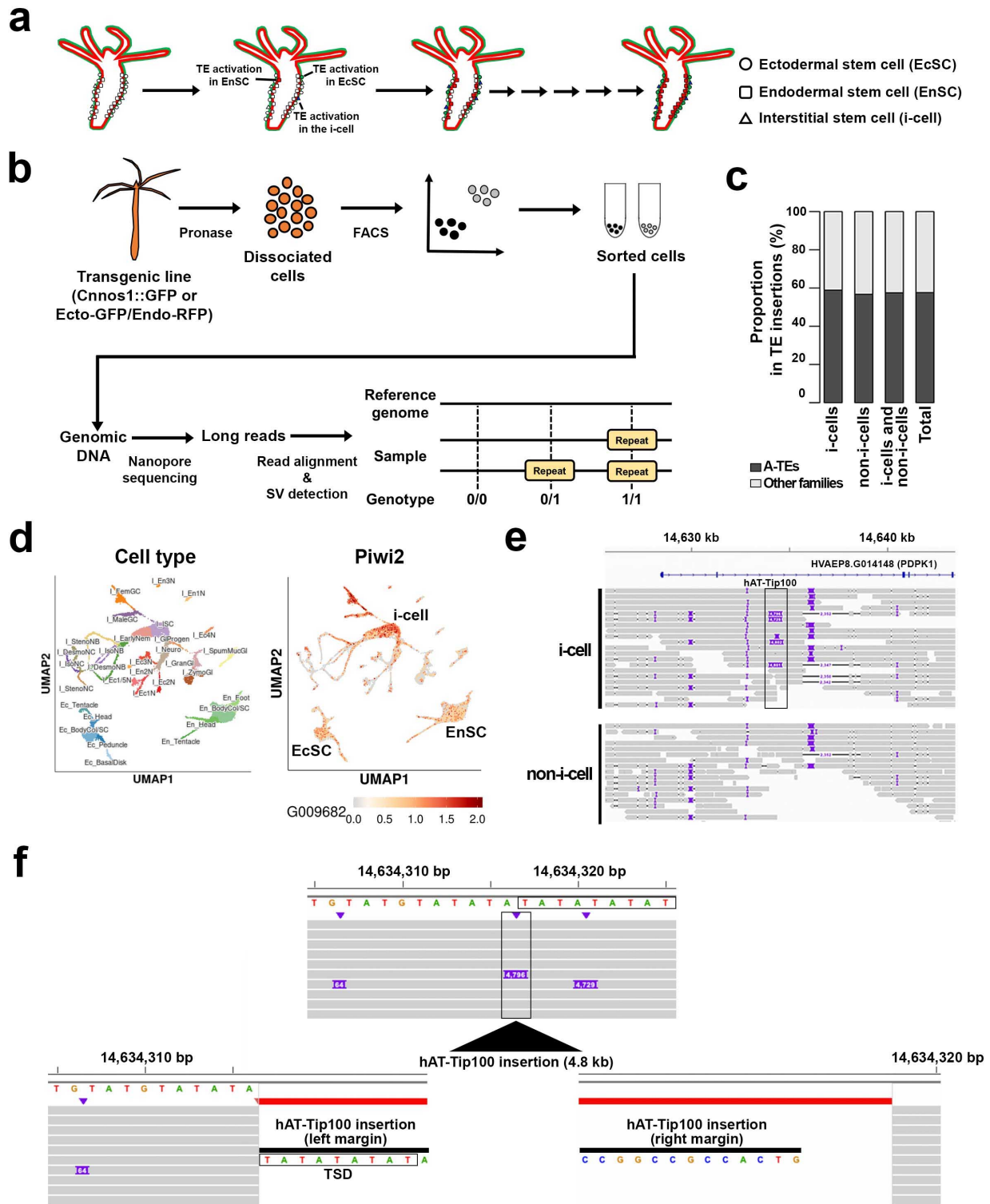


Fig. S7. Stem cell-lineage specific whole genome sequencing

(a) TE activation and stem cell genome divergence. (b) Workflow for stem cell-lineage specific whole genome sequencing. To elucidate the TE landscapes within the genomes of three stem cell lineages in hydra, transgenic lines expressing reporter genes (*Cnnos1::GFP*⁵ or *Ecto-GFP/Endo-RFP*⁶) were clonally propagated. The hydra specimens were dissociated into single cells through

Pronase treatment, and the stem cell lineages were sorted based on fluorescence intensities using FACS. Genomic DNA was then extracted from the sorted cells, and long reads were obtained via Nanopore sequencing. The reads were mapped to the AEP genome sequence to identify insertion sites. The analysis focused on insertions longer than 1kb, searching for insertions with sequences homologous to a custom repeat library generated by RepeatModeler from the genome sequence (Methods). This approach allowed for the determination of the extent to which each TE subfamily is inserted into the genome across different cell populations, facilitating the analysis of TE-induced stem cell genome divergence through comparisons across cell populations. **(c)** Contribution of the A-TEs to the TE insertions. **(d)** Piwi2 expression in the stem cell. Piwi2 (G009682) is expressed in the all three stem cell populations (i-cell, EcSC, and EnSC). Panels were retrieved from the Hydra AEP Genome Project Portal (<https://research.nhgri.nih.gov/HydraAEP/>). **(e)** The i-cell specific hAT-Tip100 insertion in the intron of 3-phosphoinositide-dependent protein kinase 1 (PDPK1) involved in cell proliferation^{7,8}. **(f)** Target site duplication of the the i-cell specific hAT-Tip100 insertion.

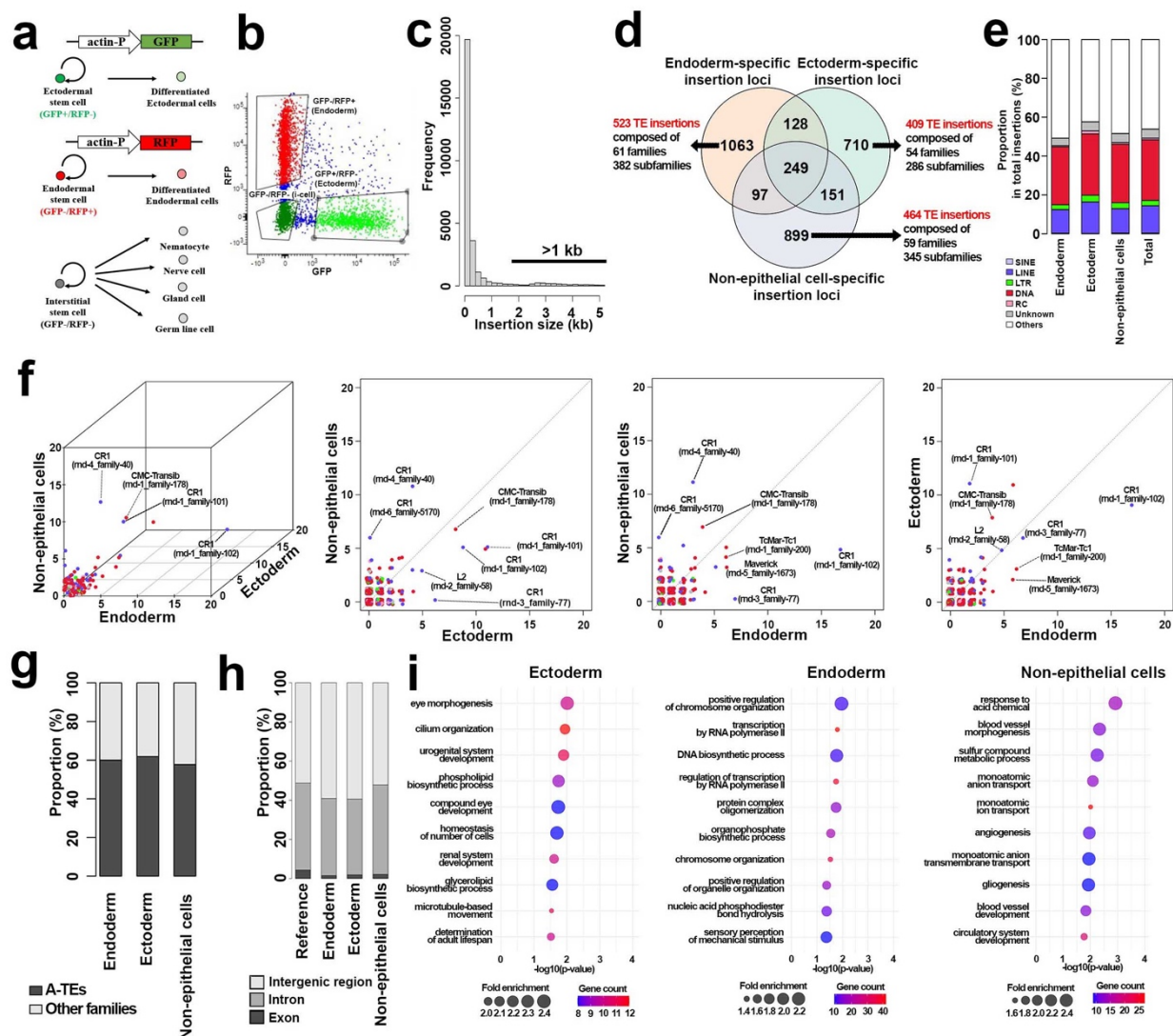


Fig. S8. TE insertion dynamics of stem cells in ecto-GFP/endo-RFP hydra

(a) ecto-GFP/endo-RFP transgenic line labelling the epithelial cells. (b) FACS-sorting of the ectodermal cell population, endodermal cell population and non-epithelial cell population. (c) Size distribution of insertions. (d) Number of cell type specific and shared insertions. (e) Major types of TEs contributing to the insertions. (f) Counts of TE subfamilies in the TE insertions at the specific loci for each population. (g) Contribution of the A-TEs to the insertions. (h) Distributions of insertions identified in intergenic regions, introns, and exons. (i) GO-term enrichment analysis of closest genes to the insertions detected.

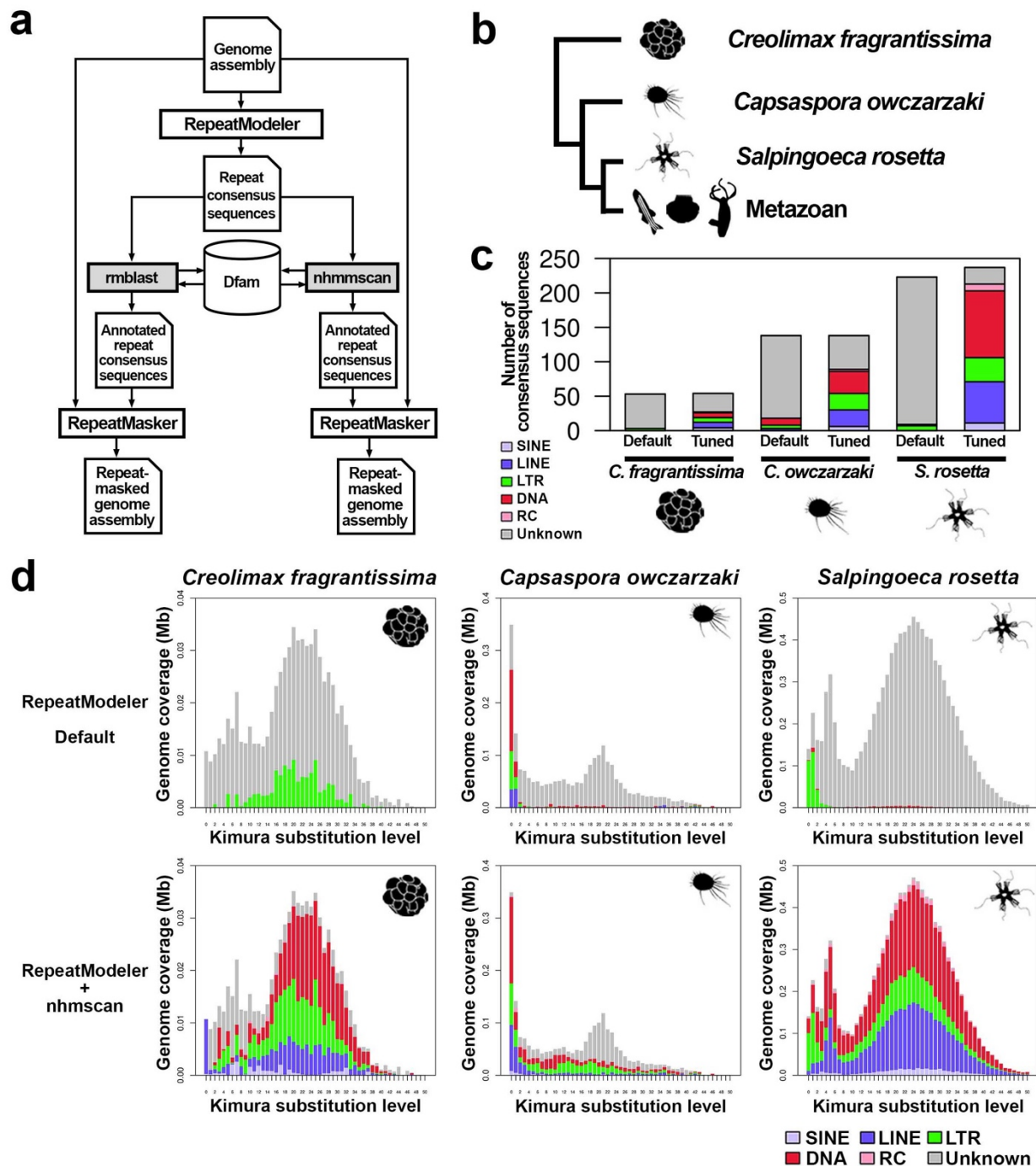


Fig. S9. Refinement of the TE detection pipeline

(a) Flowchart of the pipeline for detecting TEs. (b) Cladogram of species used for the benchmark analysis. (c) Annotation of repeat consensus sequences generated by RepeatModeler. (d) Changes in the repeat landscape.

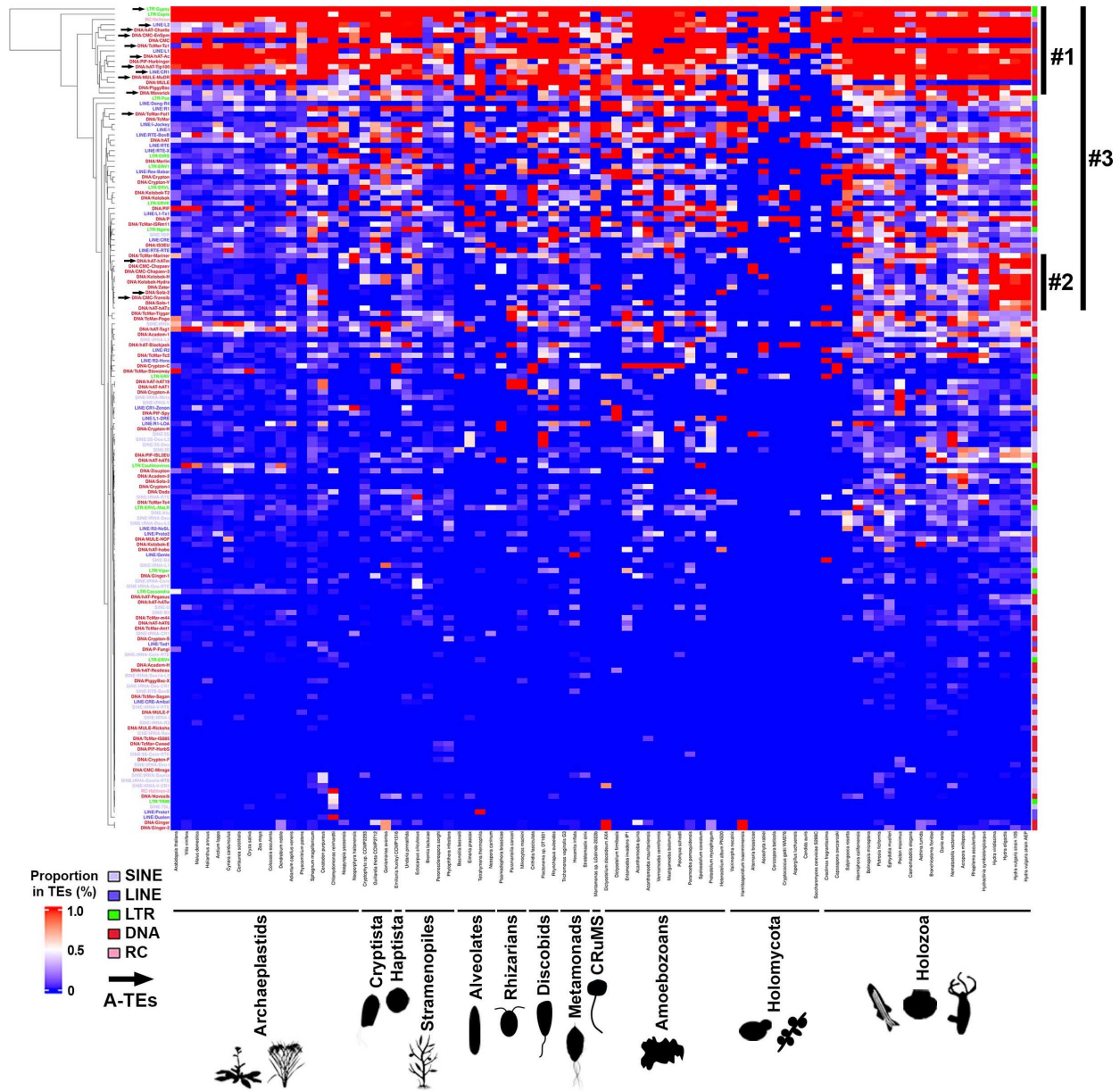


Fig. S10. Deep TE family homology across eukaryotes (related to Fig. 5)

Compositions of TEs in 82 eukaryote genomes. The heatmap represents the extent to which the genome coverage of each TE family constitutes the overall genome coverage of TEs for each species. The dendrogram at the left of the heatmap represents the result of hierarchical clustering performed using the Euclidean distance metric and the Ward's method. TEs in the cluster #3 are highlighted in Fig. 5.

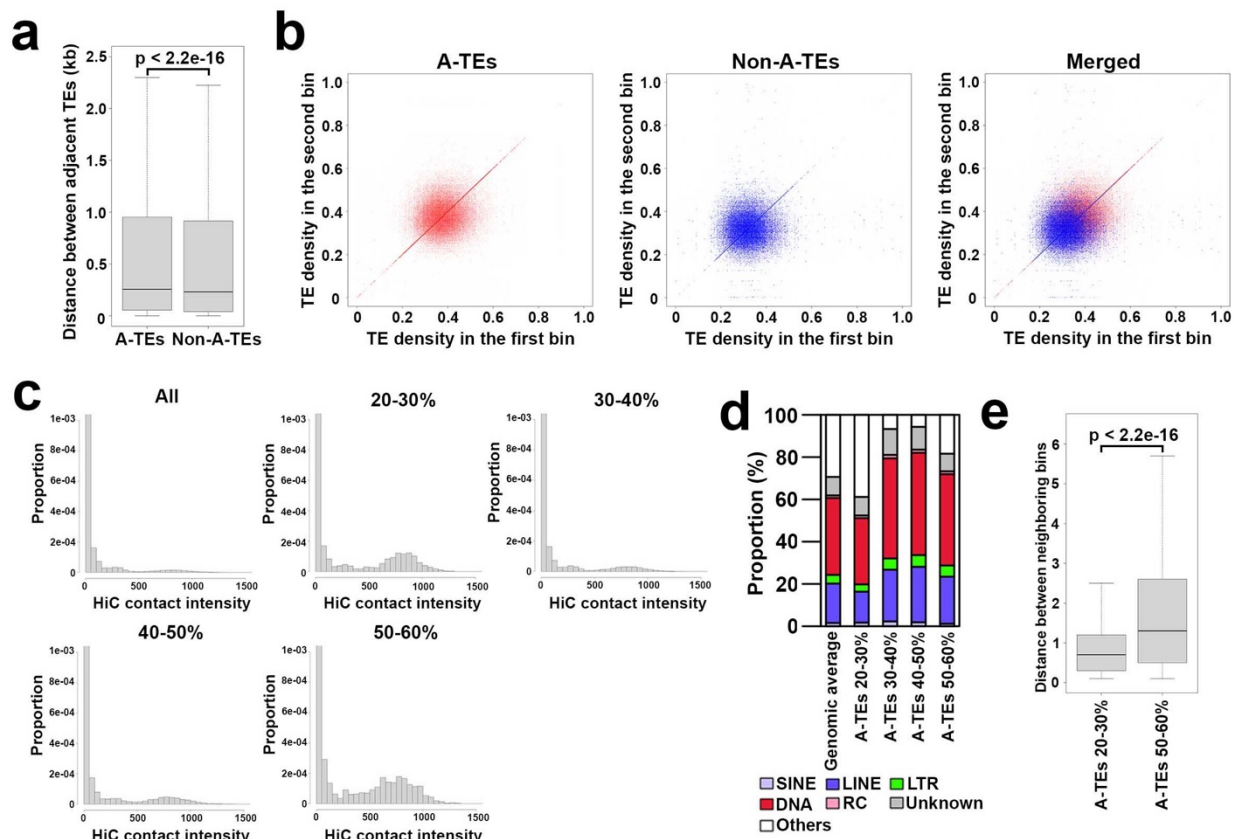


Fig. S11. A-TEs and chromosomal conformation

(a) The distance between adjacent A-TEs and between adjacent non-A-TEs. **(b)** TE density and chromatin contact intensity. Both axes represent TE density in bins, while chromatin contact intensities are indicated by the color gradients. **(c)** Relationship between A-TE coverage in bins and chromatin contact intensity. **(d)** Average TE coverage across bins. **(e)** The relationship between A-TEs coverage within a bin and the distance between bins. For bins with 20-30% A-TEs and bins with 50-60% A-TEs, the distances between bins were measured by extracting only bins with Hi-C contact intensity of 500-1000.

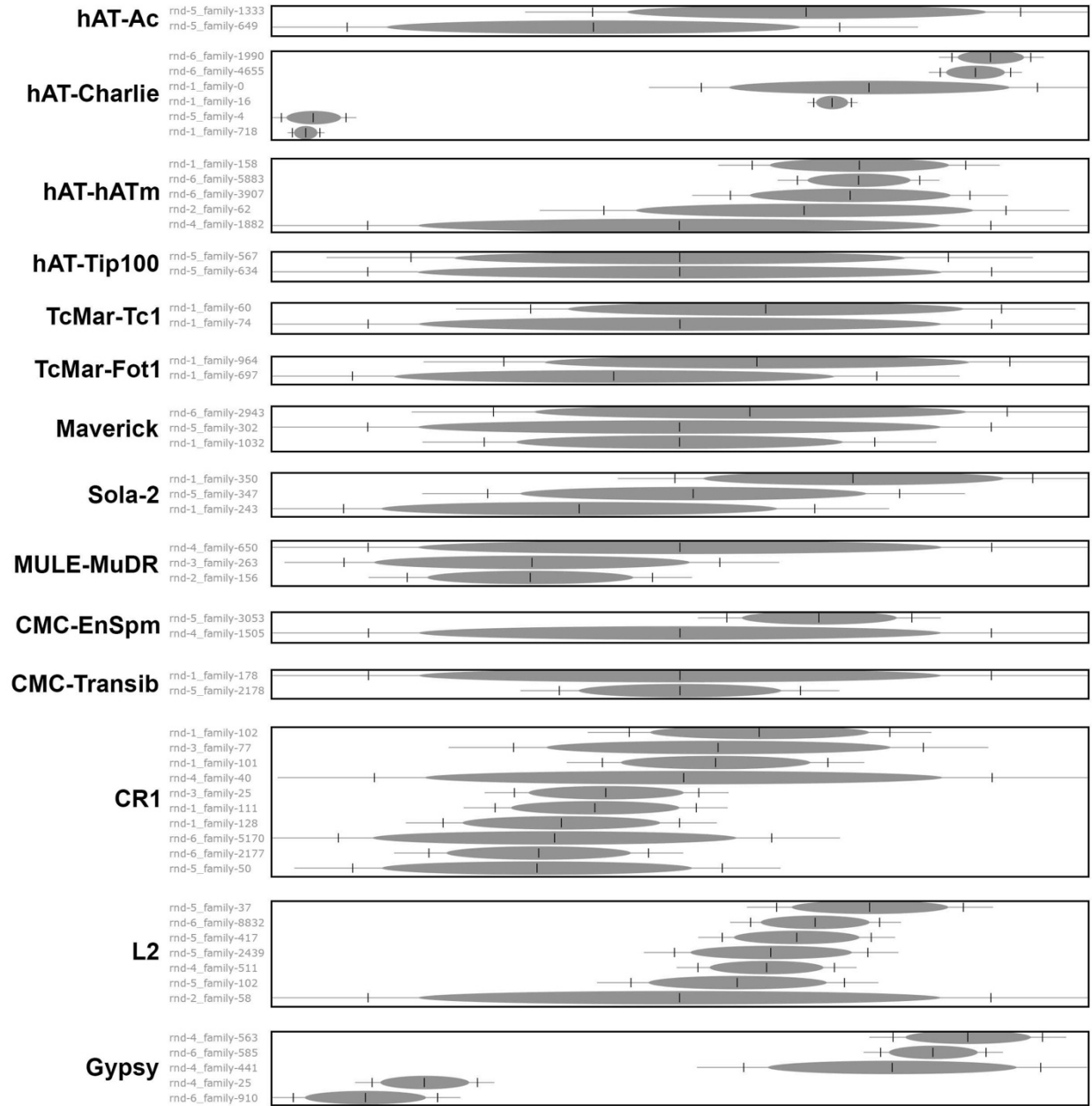


Fig. S12. Nested insertion patterns of A-TEs

Nested insertion patterns are shown for each A-TE family, and for each of its comprising RepeatModeler subfamilies (consensus sequences). Relative TinT timescale: old elements (left, diverse TinTs), young elements (right, less TinTs). The center of the ovals represents the maximum of each activity period, the end of the oval indicates 75%, and the vertical outer lines represent 99% of the probable activity period.

Supplementary Note 1. Brief summary of TE evolution and function, and the context to this study

Interspersed repetitive elements have long been studied in animal genome evolution and beyond^{9–11}. Specifically for animals, the sequencing of the first “non-model” genomes have revealed a “complex” repeat repertoire^{2,12}, comprised not only of simple repeats, but also many representatives of major DNA and Retroelement classes¹³. Each of these studies have identified hundreds to thousands of repetitive element families, many of which were found to be expanded in a species-specific pattern.

The function of these repeats has been discussed along several research lines. First, the most obvious result of the accumulation of repetitive elements is the increase in the genome size. Many studies have reported how increase in several^{14,15} or single¹⁶ repeat families or classes can result in several-fold increase in the genome size. Some of the most striking observation was the repeat expansions present in salamanders, lungfishes^{17–19}, other vertebrates²⁰, and plants²¹. Interestingly, as several studies have found, such expansions do not play a major role in large-scale genome rearrangements and the ancestral linkage groups are preserved^{1,18}. This is also evident in species that have prevailed at repetitive element expulsion, and these species often show highly rearranged genomes^{12,22,23}. Nevertheless, repeat elements, when accumulated in hotspots, can cause enhanced recombination, in particular in mammals^{24,25}.

The role of repeats in generating novel regulatory elements has been explored in many systems²⁶, most functional data available for mammals²⁷. Insertion of repeats can affect gene expression and adaptation^{28,29}, as well as drive formation of putative enhancers and other regulatory regions^{30,31}. Certain repeat classes and satellite repeats have been shown to facilitate duplications of particular gene families, such as zinc-fingers^{32,33} or long noncoding RNAs (lncRNAs)³⁴. More recently, a more direct role of protein products from certain retro-elements during fish development³⁵ or transcription factor evolution³⁶. Retro-elements have also been recently implicated in structural folding of chromatin³⁷.

The functional role of somatic insertions of repeat elements has been less investigated beyond model species. In *Drosophila*, structural variant accumulation in neuronal genomes has been debated³⁸ while general expression of repeats in various organisms has been widely reported^{25,39–41}.

These data paint a very diverse and often anecdotal picture of repeat evolution and function that, due to the large numbers of repeats and their loci and their complex evolutionary relationships, makes it a daunting task to identify general principles of TE accumulation in metazoan genomes. To this end, the biggest missing point has been the lack of a comparative and anciently branching model where TE activity can be dissected at single-cell transcriptional and genomic levels.

Supplementary Note 2. TE cross-validation and subfamily-level quantification

For cross-validation of the TE annotation in the current study, we evaluated the consistency of the TE annotation in this study with other database sources^{42,43}. We found that the masked genomic regions in the genome using the custom repeat library from the current study generally overlapped

with regions annotated as corresponding repeat families in RepBase⁴² library (Additional file 2: Table S12). For example, genomic regions marked as LTR/Gypsy superfamily using our custom repetitive element library were found to correspond to hundreds of known Gypsy families in the RepBase. However, only 8 of these families (out of 27,549 Gypsy sequences) occupied more than 50% of all Gypsy genomic regions based on the RepBase (Additional file 2: Table S13). Similarly, in the dataset of TE insertions in stem cell lineages, out of 162 Gypsy insertions (based on our Dfam-based analysis), 155 (95.7%) corresponded to 124 Gypsy sequences in Repbase, with Gypsy-19_HM-I being the most common. (Additional file 2: Table S14). Similar classification was reached using GypsyDB⁴³ as reference (Additional file 2: Table S15). These results indicate a notable consistency between the TE annotations in this study and established database resources.

In addition to the family-level quantification outlined in the main text, this section also describes subfamily (consensus sequence) level analysis. Overall similar trends were observed between family and subfamily levels. Due to their higher numbers subfamilies of each A-TE family from show a more diverse size distribution between the AEP and 105 genomes (Additional file 1: Fig. S3). For example, only few TcMar-Tc1 and CR1 subfamilies were enriched in the AEP genome while the majority of hAT-Ac and Gypsy subfamilies had a contribution to the AEP genome size increase (Additional file 1: Fig. S3).

At the subfamily level, expression of 868 subfamilies and 762 subfamilies were observed from the AEP and the 105 haplotype, respectively. TE subfamily coverage in the Iso-seq reads of the AEP strain was 59%±21 (mean±standard deviation), and that of the 105 was 57%±27 suggesting that more than half of regions of each TE-derived read is derived from a TE on average (Additional file 1: Fig. S5e). At the subfamily level, 44 types of TEs showed cell-type-specific expression profiles, and at the loci-level analysis, 203 types of TEs exhibited cell-type-specific expression profiles (Fig. 3e, Additional file 1: Fig. S6, Additional file 2: Table S6,7). The expression profiles measured by scRNA-seq were also supported by FACS-sorted bulk RNA-seq datasets⁴ (Additional file 1: Fig. S6, Methods).

In respect to insertions, 322 TE subfamilies were detected in the i-cell population; the non-i-cell population insertions was associated with 200 TE subfamilies; and the shared insertions contained 1,013 TE subfamilies (Fig. 4d). Of the TE insertions identified, 58.9% were in the i-cell population, 56.7% in the non-i-cell population, and 57.5% in those found in both the i-cell and non-i-cell populations were attributed to the A-TEs suggesting the high activities of the A-TEs in each of the stem cell lineage (Additional file 1: Fig. S7c). In both the i-cell and non-i-cell populations, the CR1 (rnd-1_family-102) and TcMar-Tc1 (rnd-1_family-200) subfamilies were the most prevalent TE subfamily insertions (Fig. 4g). In the scRNA-seq datasets of the hydra polyps³, TE subfamilies, which showed genomic insertions in the stem-cell lineage, showed relatively broad expression suggesting underlying potential contributions of chromatin states to the TE insertion preference (Fig. 4h). Furthermore, in the ecto-GFP/endo-RFP transgenic line, we identified 286 subfamilies in ectodermal TE insertions, 382 subfamilies in endodermal TE insertions, and 345 subfamilies in non-epithelial cell populations (Additional file 1: Fig. S8c,d).

References

1. Simakov, O. *et al.* Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022).
2. Chapman, J. A. *et al.* The dynamic genome of Hydra. *Nature* **464**, 592–6 (2010).
3. Siebert, S. *et al.* Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science* **365**, (2019).
4. Nishimiya-Fujisawa, C. *et al.* An ancient split of germline and somatic stem cell lineages in Hydra. *bioRxiv* (2023).
5. Nishimiya-Fujisawa, C. & Kobayashi, S. Germline stem cells and sex determination in Hydra. *Int. J. Dev. Biol.* **56**, 499–508 (2012).
6. Glauber, K. M. *et al.* A small molecule screen identifies a novel compound that induces a homeotic transformation in Hydra. *Development* **142**, 2081 (2015).
7. Gagliardi, P. A., di Blasio, L. & Primo, L. PDK1: A signaling hub for cell migration and tumor invasion. *Biochim. Biophys. Acta* **1856**, 178–88 (2015).
8. Raimondi, C. & Falasca, M. Targeting PDK1 in cancer. *Curr. Med. Chem.* **18**, 2763–9 (2011).
9. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–68 (2007).
10. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405 (2008).
11. Wells, J. N. & Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* **54**, 539–561 (2020).
12. Putnam, N. H. *et al.* Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization. *Science* (80-.). **317**, 86–94 (2007).
13. Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
14. Naville, M. *et al.* Massive Changes of Genome Size Driven by Expansions of Non-autonomous Transposable Elements. *Curr. Biol.* **29**, 1161-1168.e6 (2019).
15. Blommaert, J., Riss, S., Hecox-Lea, B., Mark Welch, D. B. & Stelzer, C. P. Small, but surprisingly repetitive genomes: transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species complex. *BMC Genomics* **20**, 466 (2019).
16. Wong, W. Y. *et al.* Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus Hydra. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22915–22917 (2019).
17. Nowoshilow, S. *et al.* The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50–55 (2018).
18. Meyer, A. *et al.* Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* **590**, 284–289 (2021).
19. Sun, C. & Mueller, R. L. Hellbender Genome Sequences Shed Light on Genomic Expansion at the Base of Crown Salamanders. *Genome Biol. Evol.* **6**, 1818–1829 (2014).
20. Kapusta, A., Suh, A. & Feschotte, C. Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci.* **114**, (2017).
21. Tenaillon, M. I., Hollister, J. D. & Gaut, B. S. A triptych of the evolution of plant transposable elements. *Trends Plant Sci.* **15**, 471–478 (2010).
22. DeBarry, J. D. & Kissinger, J. C. Jumbled Genomes: Missing Apicomplexan Synteny. *Mol. Biol. Evol.* **28**, 2855–2871 (2011).

23. Kissinger, J. C. & DeBarry, J. Genome cartography: charting the apicomplexan genome. *Trends Parasitol.* **27**, 345–354 (2011).
24. McVean, G. What drives recombination hotspots to repeat DNA in humans? *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 1213–1218 (2010).
25. O'Neill, R. J. W., O'Neill, M. J. & Graves, J. A. M. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* **393**, 68–72 (1998).
26. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017).
27. Emera, D. & Wagner, G. P. Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. *Proc. Natl. Acad. Sci.* **109**, 11246–11251 (2012).
28. Papolu, P. K. *et al.* Retrotransposons: How the continuous evolutionary front shapes plant genomes for response to heat stress. *Front. Plant Sci.* **13**, (2022).
29. Gianfrancesco, O., Bubb, V. J. & Quinn, J. P. SVA retrotransposons as potential modulators of neuropeptide gene expression. *Neuropeptides* **64**, 3–7 (2017).
30. Du, A. Y., Chobirko, J. D., Zhuo, X., Feschotte, C. & Wang, T. Regulatory transposable elements in the encyclopedia of DNA elements. *Nat. Commun.* **15**, 7594 (2024).
31. Fueyo, R., Judd, J., Feschotte, C. & Wysocka, J. Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* **23**, 481–497 (2022).
32. Wells, J. N. *et al.* Transposable elements drive the evolution of metazoan zinc finger genes. *Genome Res.* **33**, 1325–1339 (2023).
33. Eichler, E. E. *et al.* Complex β -Satellite Repeat Structures and the Expansion of the Zinc Finger Gene Cluster in 19p12. *Genome Res.* **8**, 791–808 (1998).
34. Kapusta, A. *et al.* Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).
35. Chang, N.-C. *et al.* Gag proteins encoded by endogenous retroviruses are required for zebrafish development. *bioRxiv* **10.1101/20**, (2024).
36. Cosby, R. L. *et al.* Recurrent evolution of vertebrate transcription factors by transposase capture. *Science (80-.).* **371**, (2021).
37. Lu, J. Y. *et al.* Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res.* **31**, 613–630 (2021).
38. Li, W. *et al.* Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nat. Neurosci.* **16**, 529–531 (2013).
39. Petersen, H. O. *et al.* A Comprehensive Transcriptomic and Proteomic Analysis of Hydra Head Regeneration. *Mol. Biol. Evol.* **32**, 1928–47 (2015).
40. Albertin, C. B. *et al.* The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220–224 (2015).
41. Dion-Côté, A.-M., Renaut, S., Normandeau, E. & Bernatchez, L. RNA-seq Reveals Transcriptomic Shock Involving Transposable Elements Reactivation in Hybrids of Young Lake Whitefish Species. *Mol. Biol. Evol.* **31**, 1188–1199 (2014).
42. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
43. Llorens, C. *et al.* The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* **39**, D70–D74 (2011).

Results

Paper II

Title: “Chromosome-level genome assembly of *Hydractinia symbiolongicarpus*”

Authors:

Koto Kon-Nanjo¹, Tetsuo Kon^{1*}, Helen R Horkan², Febrimarsa², Robert E Steele³, Paulyn Cartwright⁴, Uri Frank^{2*}, Oleg Simakov^{1*}

¹ Department of Neurosciences and Developmental Biology, University of Vienna, Vienna, 1030, Austria

² Centre for Chromosome Biology, School of Biological and Chemical Sciences, University of Galway, Galway, H91W2TY, Ireland

³ Department of Biological Chemistry, University of California Irvine, Irvine, CA, 92697-1700, USA

⁴ Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS, 66045, USA

*Correspondence to Oleg Simakov (oleg.simakov@univie.ac.at), Uri Frank (uri.frank@universityofgalway.ie), and Tetsuo Kon (kontetsuo@gmail.com)

Status:

Published in G3: Genes, Genomes, Genetics, 2023, 13(8):jkad107.
doi: 10.1093/g3journal/jkad107.

Chromosome-level genome assembly of *Hydractinia symbiolongicarpus*

Koto Kon-Nanjo,¹ Tetsuo Kon,^{1,*} Helen R. Horkan,² Febrimarsa,² Robert E. Steele,³ Paulyn Cartwright,⁴ Uri Frank,^{2,*} Oleg Simakov^{1,*}

¹Department of Neurosciences and Developmental Biology, University of Vienna, Vienna 1030, Austria

²Centre for Chromosome Biology, School of Biological and Chemical Sciences, University of Galway, Galway H91W2TY, Ireland

³Department of Biological Chemistry, University of California Irvine, Irvine, CA 92697-1700, USA

⁴Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045, USA

*Corresponding author: Oleg Simakov, Department of Neurosciences and Developmental Biology, University of Vienna, Vienna 1030, Austria, Email: oleg.simakov@univie.ac.at; Uri Frank, Centre for Chromosome Biology, School of Biological and Chemical Sciences, University of Galway, Galway H91W2TY, Ireland, Email: uri.frank@universityofgalway.ie; Tetsuo Kon, Department of Neurosciences and Developmental Biology, University of Vienna, Vienna 1030, Austria, Email: kontetsuo@gmail.com

Abstract

Hydractinia symbiolongicarpus is a pioneering model organism for stem cell biology, being one of only a few animals with adult pluripotent stem cells (known as i-cells). However, the unavailability of a chromosome-level genome assembly has hindered a comprehensive understanding of global gene regulatory mechanisms underlying the function and evolution of i-cells. Here, we report the first chromosome-level genome assembly of *H. symbiolongicarpus* (HSymV2.0) using PacBio HiFi long-read sequencing and Hi-C scaffolding. The final assembly is 483 Mb in total length with 15 chromosomes representing 99.8% of the assembly. Repetitive sequences were found to account for 296 Mb (61%) of the total genome; we provide evidence for at least two periods of repeat expansion in the past. A total of 25,825 protein-coding genes were predicted in this assembly, which include 93.1% of the metazoan Benchmarking Universal Single-Copy Orthologs (BUSCO) gene set. 92.8% (23,971 genes) of the predicted proteins were functionally annotated. The *H. symbiolongicarpus* genome showed a high degree of macrosynteny conservation with the *Hydra vulgaris* genome. This chromosome-level genome assembly of *H. symbiolongicarpus* will be an invaluable resource for the research community that enhances broad biological studies on this unique model organism.

Keywords: *Hydractinia*, chromosome-level genome assembly, HSymV2.0, stem cell, hydrozoa

Introduction

Hydractinia symbiolongicarpus is a colonial hydrozoan cnidarian species. The organism is distributed along the east coast of North America from Maine to South Carolina (Buss and Yund 1989). In its natural habitat, *H. symbiolongicarpus* forms colonies on hermit crab-occupied gastropod shells. In the laboratory, *H. symbiolongicarpus* can also form colonies on glass slides and be cultured in artificial seawater. A number of laboratory strains are available for this species. Colonies have functionally specialized zooids (polyps), including the feeding polyp (gastrozooid), reproductive polyp (gonozooid), and defensive polyp (dactylozooid), which are connected by the stolonal tissue, a shared gastrovascular system (Frank et al. 2020). While forming clonal colonies, *H. symbiolongicarpus* also spawns gametes daily, providing convenient access to embryogenesis. This trait facilitates gene expression manipulation using transient RNAi-mediated genetic knockdown (DuBuc et al. 2020; Quiroga-Artigas et al. 2020), CRISPR/Cas9-mediated genome editing (Gahan et al. 2016; Sanders et al. 2018), and random-integration transgenesis (Künzel et al. 2010). Various transgenic animal clones are currently available (Chrysostomou et al. 2022).

Hydractinia is the first animal in which stem and germ cells were described by August Weismann in the 19th century (Weismann 1883). The animal has a unique population of adult pluripotent stem cells, known as interstitial stem cells (i-cells for short). i-cells continuously produce progenitors to all somatic and germ cell lineages (DuBuc et al. 2020; Varley et al. 2023). A single i-cell grafted from one colony to another can self-renew and differentiate into all somatic lineages and germ cells (Varley et al. 2023). The pluripotency of *H. symbiolongicarpus* i-cells contrasts with the multipotency of *Hydra vulgaris* i-cells, which produce the neuroglandular lineage and germ cells but not epidermal and gastrodermal epithelial cells (Bosch and David 1987; Bode 1996; Hobmayer et al. 2012; Juliano et al. 2014). *H. symbiolongicarpus* is an excellent model animal for studying the function and evolution of stem cells in multicellular organisms (Plickert et al. 2012).

The genome size of *H. symbiolongicarpus* was estimated to be 514 Mb (Frank et al. 2020). *H. symbiolongicarpus* has 15 chromosomes revealed by the G-banding method (Chen et al. 2023). This chromosome number is similar in other cnidarians such as *H. vulgaris* (Simakov et al. 2022; Cazet et al. 2023). In addition, there are genetic linkage maps available, which are composed of 15 linkage groups (Chen et al. 2023). So far, the available genome sequence of

Received: March 06, 2023. Accepted: May 08, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

H. symbiolongicarpus is limited to a contig-level (>4,800 contigs) assembly called Hsym_primary_v1.0 (<https://research.nhgri.nih.gov/hydractinia/>). The unavailability of a chromosome-level genome assembly has hindered investigation into many aspects of stem cell function in *Hydractinia*, such as promoter–enhancer interactions. As such, production of a chromosome-level genome assembly of *H. symbiolongicarpus* has been anticipated for a long time by the scientific community and will provide a comprehensive landscape of the genome.

Here, we report the first chromosome-level genome assembly of *H. symbiolongicarpus* (HSymV2.0) using PacBio high-fidelity (HiFi) sequencing and Hi-C scaffolding. The final assembly is 483 Mb in total length with 15 chromosomes, representing 99.8% of the total assembly size. It will be an invaluable resource for studying molecular regulatory mechanisms underlying development, stem cells, regeneration, and their evolution.

Materials and methods

Animal husbandry

Colonies of *H. symbiolongicarpus* were maintained as described previously (Frank et al. 2020). The animals were grown on glass slides kept in artificial seawater at 20–22°C. The animals were kept in a constant 14:10 light:dark cycle.

DNA sequencing library preparation and sequencing

High-molecular weight genomic DNA was extracted from male feeding polyps (clone 291-10) of *H. symbiolongicarpus*. A PacBio HiFi library was prepared according to the manufacturer's protocol and sequenced using the PacBio Sequel II platform. For Hi-C library preparation, male (clone 291-10) feeding polyps of *H. symbiolongicarpus* were flash frozen and ground. Following the manufacturer's instructions, a Hi-C library was prepared using the Dovetail Omni-C kit (Canta Bio, #21005). The insert size distribution of the Hi-C library was confirmed to be approximately between 350 bp and 1,000 bp using Agilent Fragment Analyzer (Agilent). The Hi-C library was sequenced with 150 bp paired-end reads using Illumina MiSeq by the Next Generation Sequencing core facility of the Vienna BioCenter.

Genome assembly and scaffolding

Statistics of the PacBio HiFi read quality were calculated using NanoPlot v1.41.0 (De Coster et al. 2018). PacBio HiFi reads were assembled into contig sequences using Hifiasm v0.16.1-r375 (Cheng et al. 2021) with default parameters. The chromosome-level scaffolds were constructed with the combination of the contig sequences and the Hi-C sequencing reads. The Hi-C reads were aligned to the contig sequences using Juicer v1.6 (Durand et al. 2016b) with default parameters. Then, Hi-C scaffolding was performed using the 3D-DNA pipeline (Dudchenko et al. 2017) with parameters “-editor-repeat-coverage 1000 -r 0.” Scaffolds were manually curated using Juicebox Assembly Tools v1.11.08 (Durand et al. 2016a). The final assembly was named HSymV2.0. Assembly quality and completeness were evaluated using QUAST v5.2.0 (Gurevich et al. 2013) and Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.2.2 (Simão et al. 2015) with the metazoa odb10 database (metazoa_odb10).

Validation of the genome assembly with *H. symbiolongicarpus* linkage maps

In order to anchor the contig sequences of the Hsym_primary_v1.0 assembly to the previously reported genetic linkage maps of *H.*

symbiolongicarpus (Chen et al. 2023), we merged the contig sequences based on the SNP marker positions and constructed 15 pseudochromosome sequences for the maternal and paternal genetic linkage map. These pseudochromosome sequences were aligned to the 15 chromosomal-level scaffolds of the HSymV2.0 assembly using minimap2 v2.24 (Li 2018) with the parameters “-ax asm5 -cs -secondary=no.” Alignments with mapping quality 60 were retained using samtools view (Danecek et al. 2021). The alignment files in the BAM format were converted to the PAF format using paftools from minimap2 v2.24. Finally, the alignments were visualized in dot plots using D-GENIES (Cabanettes and Klopp 2018).

Identification of repetitive sequences

A custom repeat library of the genome assembly was generated using RepeatModeler v2.0.4 (Flynn et al. 2020) with default parameters. Next, repetitive sequences were identified using RepeatMasker v4.1.4 (<http://www.repeatmasker.org>) and the custom repeat library. A repeat-masked sequence of the HSymV2.0 assembly was also generated at this step. Finally, the output of RepeatMasker was summarized using the helper scripts from RepeatMasker v4.1.4. For comparison, repetitive sequences in the *H. vulgaris* strain AEP (Cazet et al. 2023) were identified using the same pipeline.

Genome annotation

The repeat-masked genome assembly of the HSymV2.0 assembly was used for gene predictions. We identified the protein-coding region and untranslated region (UTR) for each gene, with a combination of ab initio prediction, homology-based prediction, and transcriptome-based prediction using BRAKER2 (Brüna et al. 2021). For homology-based prediction, we obtained the genome sequence and annotation of *H. vulgaris* strain 105 (GCA_022113875.1) from the National Center for Biotechnology Information (NCBI) Genome database; the genome sequence and annotation of *H. vulgaris* strain AEP (Cazet et al. 2023) from the Hydra AEP Genome Project Portal (<https://research.nhgri.nih.gov/HydraAEP/>); and the genome sequence and annotation from the contig-level reference assembly of *H. symbiolongicarpus* (Hsym_primary_v1.0) from the Hydractinia Genome Project Portal (<https://research.nhgri.nih.gov/hydractinia/>). Protein sequences were generated from the reference genome sequences and the genome annotations using GffRead (Pertea and Pertea 2020) with the parameter “-S -J.” For each gene, the longest isoform was extracted for homology-based prediction. Then, all protein sequences were merged. BRAKER2 was run on the repeat-masked genome assembly using the homologous protein sequences with parameters “-epmode -softmasking.” For transcriptome-based prediction, we used publicly available RNA-seq reads (accession numbers: SRR1796501–SRR1796515, SRR9331388–SRR9331403, and SRR14265606–SRR14265626, SRR18686538–SRR18686548). The repeat-masked genome assembly was indexed using STAR v2.7.10b (Dobin et al. 2013) with the parameter “-runMode genomeGenerate.” The RNA-seq reads were aligned to the indexed genome assembly using STAR v2.7.10b with parameters “-runMode alignReads -outSAMtype BAM SortedByCoordinate.” All alignment files were merged using samtools merge (Danecek et al. 2021). Then, BRAKER2 was run on the repeat-masked genome assembly using the merged bam file with parameters “-softmasking.” UTR regions were added to the transcriptome-based predictions using BRAKER2 with parameters “-softmasking -addUTR=on.” From the initial gene model predictions, consensus gene models were produced using TSEBRA v1.0.3 (Gabriel et al. 2021). We also transferred the gene annotation gene coordinates from the Hsym_primary_v1.0 assembly to the HSymV2.0 assembly using the Liftoff tool (Shumate and Salzberg 2020). We selected

gene models that contained start and stop codons, while excluding those that had internal stop codons. Finally, gene model predictions from BRAKER2 and gene models from Liftoff were integrated using BEDTools (Quinlan and Hall 2010). To validate the completeness of the final gene models, protein sequences were generated from the HSymV2.0 assembly and the genome annotations using GffRead (Pertea and Pertea 2020). Each protein sequence was examined for the presence of homologous sequences in the NCBI-nr database using the BLAST+ v2.13.0 with the parameters “-evalue 1e-05.” Protein domain prediction was performed using InterProScan v5.60-92.0 (Zdobnov and Apweiler 2001). Gene functional annotation was performed using eggNOG-Mapper v2.1.9 (Huerta-Cepas et al. 2019). Mitochondrial genome annotation was performed using the MITOS WebServer (Bernt et al. 2013). For visualization of the mitochondrial genome annotation, a Genbank file was generated from the mitochondrial genome sequence and its annotation using the EMBOSS Seqret tool (Rice et al. 2000). The mitochondrial genome map was generated from the Genbank file using ApE (Davis and Jorgensen 2022).

Synteny analysis

To identify orthologs between *H. symbiolongicarpus* and *H. vulgaris*, the longest protein isoforms from the chromosome-level scaffolds were prepared for both species. Both gene sets were compared reciprocally using blastp with default parameters (Altschul et al. 1990). Gene pairs with reciprocal best hits were identified as orthologs between the two species. For each species, each ortholog was numbered according to its genomic coordinate. Oxford plots were plotted using the plot function in R v4.2.2.

Results and discussion

Chromosome-level genome assembly of *H. symbiolongicarpus* (HSymV2.0)

High-molecular weight genomic DNA was extracted from male (clone 291-10) colonies of *H. symbiolongicarpus* (Fig. 1a and

Supplementary Fig. 1a), the same genetic individual from which the first version of the genome was prepared. A total of 34.0 Gb HiFi reads (66x) of the *H. symbiolongicarpus* genome were generated with a read N50 length of 13.2 kb (Table 1 and Supplementary Fig. 1b). We performed de novo genome assembly of the *H. symbiolongicarpus* genome using Hifiasm (Cheng et al. 2021). The initial genome assembly contains 325 contigs with a total length of 483 Mb (Table 2). The total contig length is comparable to the previously estimated genome size (Frank et al. 2020). The contig N50 was 28.8 Mb and the contig L50 was 8 (Table 2). The maximum contig length was 42.6 Mb (Table 2). The GC content was 35.85% (Table 2). We also assembled a 15,471 bp mitochondrial DNA scaffold which contains 13 proteins, small and large subunit rRNAs, and methionine and tryptophan tRNAs like other cnidaria species (Supplementary Fig. 2) (Zou et al. 2012; Lin et al. 2019). We evaluated the completeness of the genome assembly by searching for sets of BUSCO in the contig sequences (Simão et al. 2015). We used the metazoa odb10 database (metazoa_odb10) for this analysis. We identified 818 complete BUSCO genes out of 954 BUSCO genes (85.7%), of which 756 were present as single copies (79.2%) and 62 as duplicates (6.5%). We also identified 64 fragmented BUSCO genes (6.7%). Of note, these statistics of the contig sequences are better than those of the Hsym_primary_v1.0 assembly, the previously reported contig-level assembly (Table 3). For example, contig N50 of the current assembly is 13 times higher than that of Hsym_primary_v1.0 (28.8 Mb vs 2.2

Table 1. Summary of the genome sequencing.

Experiment	Sequencing platform	Number of reads	Total base (base)	Read length (bp)
Whole genome sequencing	PacBio Sequel II	3,049,931	34,014,052,488	13,241 (N50)
Hi-C	Illumina MiSeq	14,836,376	4,450,912,800	150 × 2

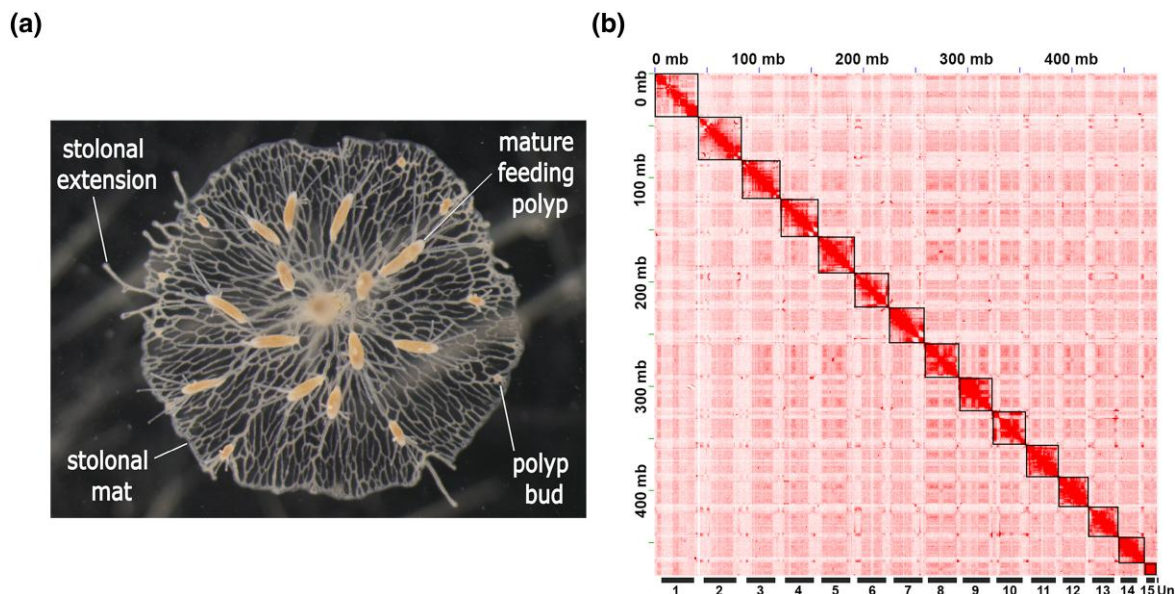


Fig. 1. Chromosome-level genome assembly of *H. symbiolongicarpus* (HSymV2.0). a) Colony of *H. symbiolongicarpus* male (Clone 291-10) attached to a glass microscope slide, showing mature feeding polyps (~5 mm height), polyp buds, and shared stolon network (sexual polyps not pictured as they develop later). b) Hi-C contact map showing the 15 chromosome-level scaffolds and the unplaced sequences (Un). Scaffolds are ordered by length and assigned numbers.

Table 2. Assembly statistics of the *H. symbiolongicarpus* genome.

	Contigs	All Hi-C scaffolds (HSymV2.0)	Chromosome-level scaffolds	Hsym_primary_v1.0
Total length (bp)	482,771,323	482,928,323	481,835,061	406,663,980
Total sequence number	325	84	15	4,840
Maximum sequence length (bp)	42,595,244	42,572,105	42,572,105	11,665,964
N50 (bp)	28,808,060	33,314,500	33,314,500	2,235,630
L50	8	7	7	48
GC content (%)	35.85	35.85	35.85	35.22

Table 3. BUSCO scores from the genome assembly.

	Current study (HSymV2.0)	Hsym_primary_v1.0
Complete BUSCOs (C)	818 (85.7%)	796 (83.5%)
Complete and single-copy BUSCOs (S)	756 (79.2%)	722 (75.7%)
Complete and duplicated BUSCOs (D)	62 (6.5%)	74 (7.8%)
Fragmented BUSCOs (F)	64 (6.7%)	69 (7.2%)
Missing BUSCOs (M)	72 (7.6%)	89 (9.3%)
Total BUSCO groups searched	954	954

Mb), and the number of contigs is less than a tenth of that of Hsym_primary_v1.0 (325 vs 4,840, Table 2). To construct a chromosome-level genome assembly, 14,836,376 Hi-C read pairs were generated from male (clone 291-10) feeding polyps (Supplementary Fig. 3). Hi-C scaffolding generated the final assembly of 483 Mb in length with 15 chromosomes which represent 99.8% of the total assembly size (Fig. 1b and Supplementary Fig. 4). The scaffold N50 was 33 Mb. The size of the largest scaffold was 42.6 Mb. The total size of the 69 unplaced sequences was 1.1 Mb. Four of the 15 scaffolds (scaffold 8, 11, 12, and 15) consist of single contigs (Supplementary Fig. 4). The Hi-C contact map of the chromosome-level genome assembly shows a low number of trans-chromosomal interactions in our genome assembly (Fig. 1b). In each chromosome, the Hi-C contact intensity at the diagonal position is stronger than that of the other position. This indicates that adjacent sequences of our assembly have strong contact. These findings suggest the completeness of the chromosome-level genome assembly. The final chromosome-level genome assembly was named HSymV2.0.

Comparing HSymV2.0 with the preexisting genetic linkage maps and sex chromosome identification

In order to further evaluate the completeness of the HSymV2.0 assembly, we compared our genome assembly with the previously reported genetic linkage maps of *H. symbiolongicarpus* (Chen et al. 2023). Based on the loci of the SNP markers in the Hsym_primary_v1.0 assembly, we anchored the contig sequences of the Hsym_primary_v1.0 to the maternal and paternal genetic linkage maps. We obtained the maternal and paternal 15 pseudochromosome sequences. Then, we aligned the maternal and paternal pseudochromosomes to our 15 chromosome-level scaffolds. Both sets of 15 pseudochromosomes were uniquely mapped to our 15 chromosome-level scaffolds (Fig. 2a and b). The resulting alignments also revealed the colinear relationships between the 15 chromosome-level scaffolds of the HSymV2.0 assembly and the maternal pseudochromosomes (Fig. 2a) or paternal pseudochromosomes (Fig. 2b). These results demonstrate the contiguity

and completeness of the HSymV2.0 assembly. The previous study showed that *H. symbiolongicarpus* has an XX/XY sex determination system and that the sex determining locus is located at the end of linkage group 4 (Chen et al. 2023). In both linkage groups, linkage group 4 was mapped to chromosome 1 of the HSymV2.0 assembly (Fig. 2a and b). Based on this result, the sex chromosome in the HSymV2.0 assembly is most likely to be chromosome 1 (Fig. 2).

Identification of repetitive sequences

For the identification of repetitive sequences in the genome assembly, we generated a custom repeat library of our genome assembly using RepeatModeler. This library contains 2,300 repeat families. Then, RepeatMasker was run on the assembly using the custom library. Repetitive sequences accounted for 296 Mb (61%) of the HSymV2.0 assembly. The repetitive sequences are composed of total interspersed repeats (58.25%), simple repeats (0.70%), and others (Table 4). The total interspersed repeats can be further divided into retroelements (5.20%), DNA transposons (4.94%), rolling circles (0.55%), and unclassified elements (48.11%). Among the retroelements, LINEs (3.01%) were found to be the most abundant (Table 4). Among the DNA transposons, Tc1-IS630-Pogo (0.48%) and hobo-Activator (0.33%) were the top two elements (Table 4). These results are similar to the repeat analysis of the *H. vulgaris* genome. In the *H. symbiolongicarpus* genome, sequence divergence of the repeats revealed at least two discrete periods of repeat expansion (Fig. 3a). In contrast, in the *H. vulgaris* genome, more continuous repeat expansion was shown (Fig. 3b), which was also indicated by the previous study (Chapman et al. 2010).

Gene prediction and functional annotation

In order to predict protein-coding genes in the HSymV2.0 assembly, we adopted a combination of ab initio prediction: homology-based prediction, and transcriptome-based prediction. For homology-based prediction, we used protein sequences from *H. vulgaris* and Hsym_primary_v1.0. For transcriptome-based prediction, publicly available RNA-seq reads were aligned to the assembly using the STAR aligner (Dobin et al. 2013). We then performed genome annotation using BRAKER2 (Brůna et al. 2021) with homology evidence or transcriptome evidence. Finally, the resulting genome annotations were integrated using TSEBRA (Gabriel et al. 2021) to produce the consensus gene model. We also mapped the gene coordinates of the gene annotation of the Hsym_primary_v1.0 assembly to the HSymV2.0 assembly using the Liftoff tool (Shumate and Salzberg 2020). Finally, we predicted a total of 25,825 protein-coding genes. Of the 22,022 protein-coding genes in the Hsym_primary_v1.0 assembly, 21,686 genes (98.5%) were mapped to 18,606 genes in the HSymV2.0 assembly (Supplementary Table 1). The predicted protein-coding genes showed a mean gene body length of 5,854 bp, a mean coding sequence (CDS) length of 1,346 bp, and a mean number of exons of 6.43. We detected 888 complete BUSCO genes out of a total of 954 BUSCO genes (93.1%) in the 25,825 predicted proteins. Among these genes, 813 appeared in single copies

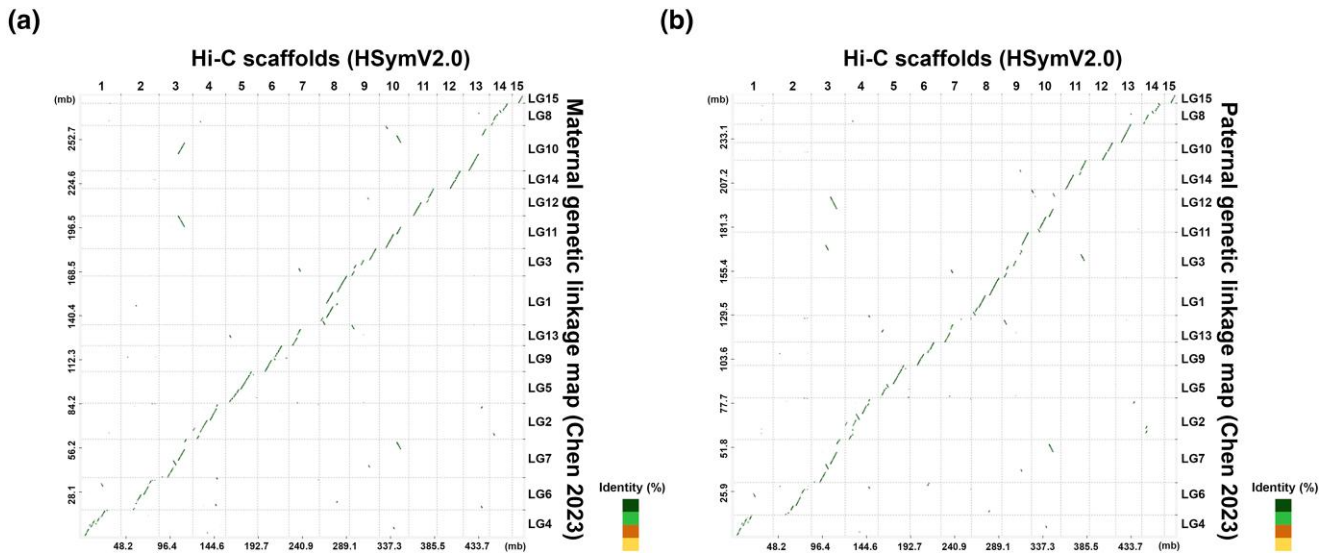


Fig. 2. Collinearity between the Hi-C scaffolds of the present study and the genetic linkage maps (Chen et al. 2023). a) Sequence alignments between the Hi-C scaffolds and pseudochromosomes based on the maternal genetic linkage map. b) Sequence alignments between the Hi-C scaffolds and pseudochromosomes based on the paternal genetic linkage map.

Table 4. Summary of the RepeatMasker outputs.

	<i>H. symbiolongicarpus</i>			<i>H. vulgaris</i>		
	Number of elements	Length occupied (bp)	Percentage of sequence	Number of elements	Length occupied (bp)	Percentage of sequence
Retroelements	50,329	25,095,050	5.20	316,787	153,628,253	17.05
SINEs:	12,586	1,987,472	0.41	21,296	3,637,401	0.40
Penelope	9,072	2,726,592	0.56	52,322	17,817,731	1.98
LINEs:	27,013	14,541,672	3.01	281,818	140,096,759	15.55
CRE/SLACS	1,786	821,583	0.17	298	150,224	0.02
L2/CR1/Rex	9,860	6,205,894	1.29	207,793	111,582,339	12.39
R1/LOA/Jockey	0	0	0.00	659	166,379	0.02
R2/R4/NeSL	1,415	881,271	0.18	0	0	0.00
RTE/Bov-B	3,902	3,319,744	0.69	3,759	1,136,092	0.13
L1/CIN4	128	79,792	0.02	0	0	0.00
LTR elements:	10,730	8,565,906	1.77	13,673	9,894,093	1.10
BEL/Pao	1,500	1,411,183	0.29	2,320	2,078,705	0.23
Ty1/Copia	2,594	1,823,078	0.38	135	160,597	0.02
Gypsy/DIRS1	5,721	4,579,944	0.95	9,802	6,193,847	0.69
Retroviral	155	104,124	0.02	0	0	0.00
DNA transposons	56,331	23,868,445	4.94	536,767	243,675,039	27.05
hobo-Activator	3,454	1,578,931	0.33	183,730	81,989,900	9.10
Tc1-IS630-Pogo	11,142	2,305,486	0.48	128,135	56,913,582	6.32
En-Spm	0	0	0.00	0	0	0.00
PiggyBac	338	264,413	0.05	4,776	2,483,575	0.28
Tourist/Harbinger	5,311	1,339,873	0.28	6,623	1,752,620	0.19
Other	3,354	1,092,285	0.23	30,175	13,042,137	1.45
Rolling-circles	9,783	2,673,725	0.55	6,653	3,214,957	0.36
Unclassified:	410,631	232,360,741	48.11	764,256	184,468,711	20.48
Total interspersed repeats		281,324,236	58.25		581,772,003	64.57
Small RNA:	12,289	8,667,551	1.79	13,979	1,876,864	0.21
Satellites:	918	166,061	0.03	0	0	0.00
Simple repeats:	51,000	3,398,212	0.70	376,701	41,350,470	4.59
Low complexity:	10,920	531,858	0.11	49,800	2,588,884	0.29
Total bases masked:		295,609,498	61.21		630,803,178	70.02

(85.2%), while 75 existed as duplicates (7.9%) (Table 5). These BUSCO scores were better than those of the predicted protein sequences of the Hsym_primary_v1.0 assembly (Table 5). Of all predicted protein-coding genes in our assembly, 20,985 protein-coding genes (81.3%) had significant homology to the sequences in the NCBI

nonredundant database (nr) (Fig. 4a and Supplementary Table 2). Out of the total 20,985 genes that exhibited significant homology to genes in the NCBI-nr database, 16,670 genes (79.4%) were hit to Cnidaria genes (Fig. 4a). 4,255 genes (20.3%) were hit to Bilateria genes (Fig. 4a). The remaining 42 (0.2%) and 18 (0.1%) genes hit the

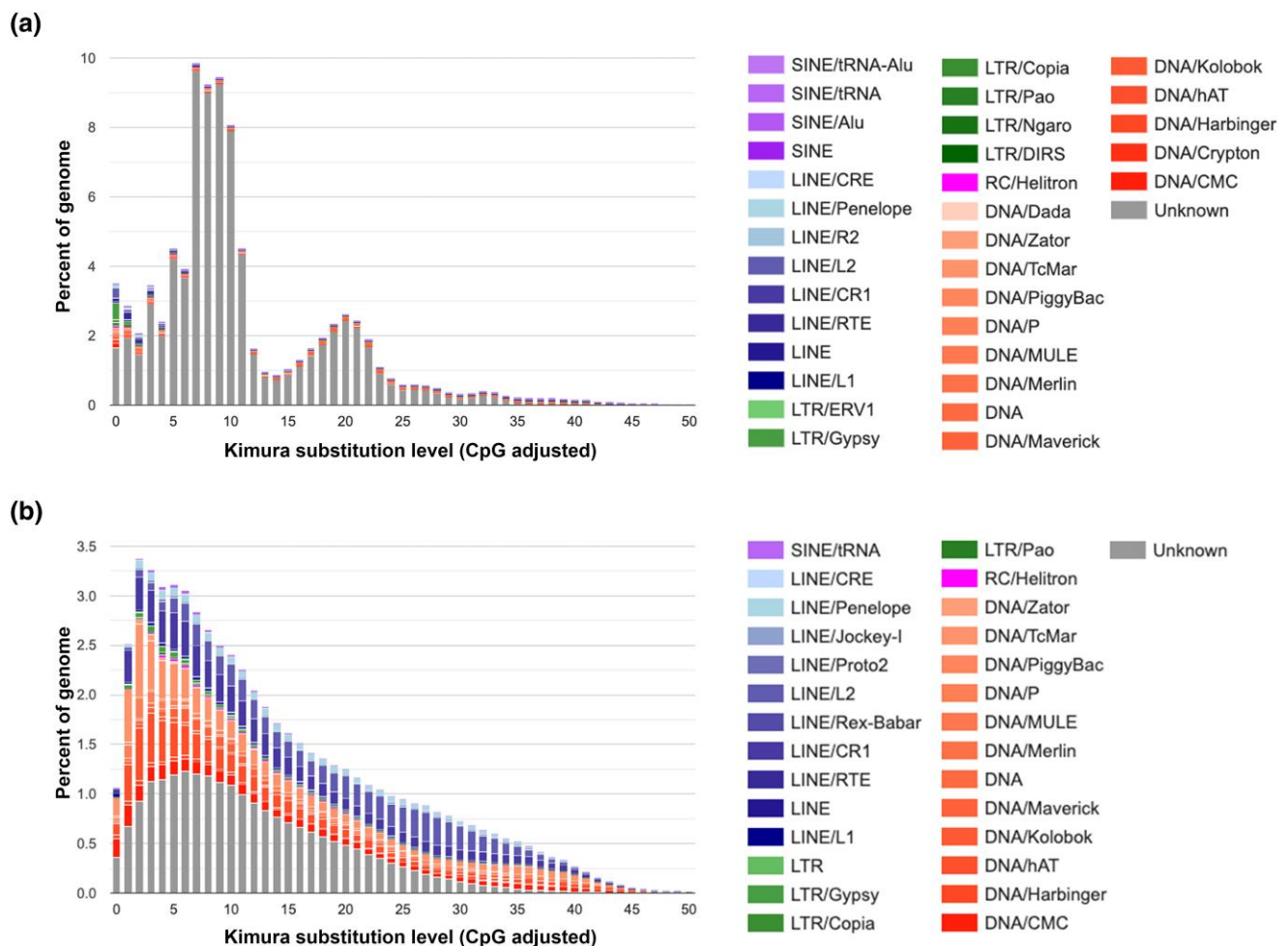


Fig. 3. Interspersed repeat landscape of the *H. symbiolongicarpus* genome and *H. vulgaris* genome. a) Interspersed repeat landscape of *H. symbiolongicarpus*. b) Interspersed repeat landscape of *H. vulgaris*. The x-axis represents the level of Kimura substitution for repeat elements from the consensus sequences (relative age). The y-axis represents the relative abundance of each repeat family in the genome. Ancient active repeats are placed on the right side of the graph, and recently active repeats are on the left.

Table 5. BUSCO scores from the predicted proteins.

	Current study (HSymV2.0)	Hsym_primary_v1.0
Complete BUSCOs (C)	888 (93.1%)	882 (92.5%)
Complete and single-copy BUSCOs (S)	813 (85.2%)	785 (82.3%)
Complete and duplicated BUSCOs (D)	75 (7.9%)	97 (10.2%)
Fragmented BUSCOs (F)	20 (2.1%)	12 (1.3%)
Missing BUSCOs (M)	46 (4.8%)	60 (6.2%)
Total BUSCO groups searched	954	954

Sponge and Placozoa genes, respectively. We identified that 22,199 protein-coding genes (86.0%) had protein domains using InterProScan (Supplementary Table 3) (Zdobnov and Apweiler 2001). We also used eggNOG-Mapper (Huerta-Cepas et al. 2019) for the predicted protein-coding genes and found that 18,547 genes were functionally annotated (Supplementary Table 4). Overall, 92.8% (23,971 genes) of all predicted protein-coding genes had functional annotations supported by one or more of the above databases (Fig. 4b). Stem cell factor genes are well predicted in our genome

annotation. For example, *Piwi1* is a known marker for i-cells (Bradshaw et al. 2015; Gahan et al. 2016). In the HSymV2.0 assembly, *Piwi1* is on the plus strand of DNA in chromosome 13, spanning from base pair position 16,493,941 to 16,509,130 (Fig. 4c). Another example is transcription factor *AP2* (*Tfap2*). This gene is a regulator of germ cell commitment (DuBuc et al. 2020). *Tfap2* is located on chromosome 5 from base pair position 20,139,307 to 20,141,788 on the plus strand of DNA (Fig. 4d).

Syntenic analysis between *H. symbiolongicarpus* and *H. vulgaris*

We investigated the syntenic relationship between *H. symbiolongicarpus* and *H. vulgaris*. A total of 8,974 orthologs were identified (Supplementary Table 5). We generated the Oxford dot plot between *H. symbiolongicarpus* and *H. vulgaris* with these orthologs (Fig. 5a). Overall, this plot revealed that chromosome-scale synteny (macrosynteny) is highly conserved without collinearity between these species (Fig. 5a). Also, pairs of chromosomes (scaffold 5 and 15 of *H. symbiolongicarpus* vs HVAEP3 and HVAEP5 of the *H. vulgaris*, scaffold 2 and 14 of *H. symbiolongicarpus* vs HVAEP7 and HVAEP14 of the *H. vulgaris*) showed syntenic relationships suggesting the past interchromosomal translocations after the divergence between the species (Fig. 5b and c).

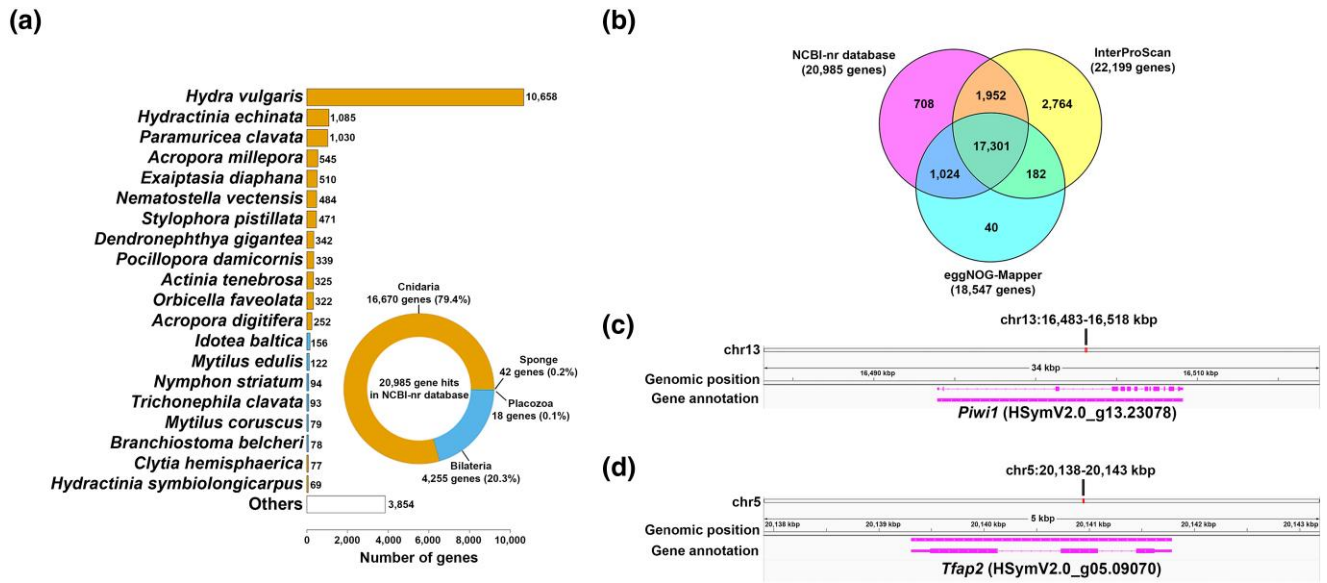


Fig. 4. Gene annotation of the HSymV2.0 assembly. a) BLAST search results of the total 25,825 predicted protein-coding genes against the NCBI-nr database. Total 20,985 gene hits were identified in the NCBI-nr database. Top 20 species origins of the BLAST hits are shown as a bar plot. The inset pie chart represents the summary of species origins. The color scheme of the bar plot corresponds to that of the inset pie chart. b) Venn diagram showing functional annotations of the predicted protein-coding genes of the HSymV2.0 assembly. c) Gene structure of *Piwi1* on the chromosome 13. d) Gene structure of *Tfap2* on the chromosome 5.

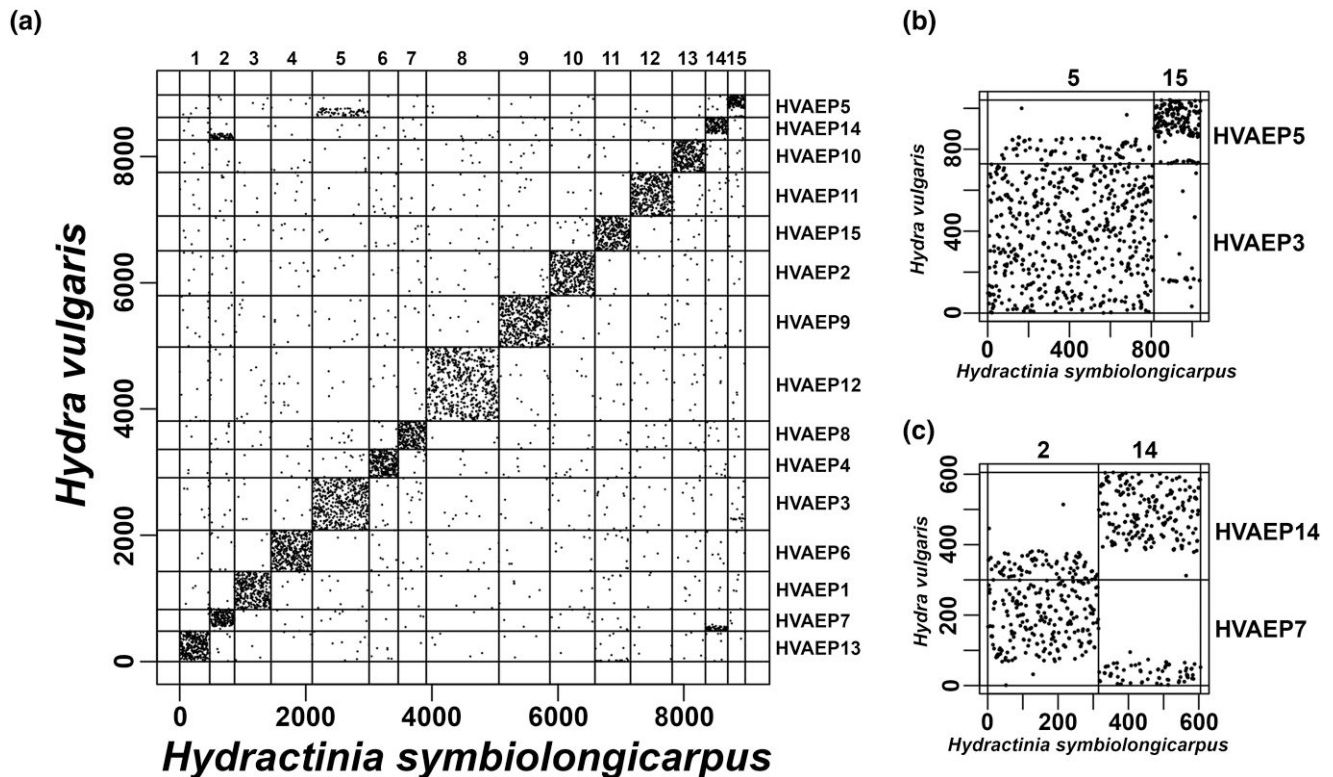


Fig. 5. Oxford dot plot of orthologs between *H. symbiolongicarpus* and *H. vulgaris*. a) Oxford dot plot using the 8,974 orthologs. For both species, orthologs are sequentially numbered according to their genomic positions. Each dot represents the orthologs. b) Oxford dot plot using orthologs on scaffold 5 and 15 of *H. symbiolongicarpus* genome assembly and those on the HVAEP3 and HVAEP5 of the *H. vulgaris* genome assembly. c) Oxford dot plot using orthologs on the scaffold 2 and 14 of *H. symbiolongicarpus* genome assembly and those on the HVAEP7 and HVAEP14 of the *H. vulgaris* genome assembly.

Conclusion

In summary, we performed de novo genome assembly of *H. symbiolongicarpus* and obtained contig sequences with a size of 483

Mb and N50 of 28.8 Mb. The following Hi-C scaffolding resulted in 15 chromosome-level scaffolds containing 99.8% of the total genome assembly. This HSymV2.0 assembly is the first chromosome-level genome assembly of *H. symbiolongicarpus*. In

the HSymV2.0 assembly, we predicted 25,825 protein-coding genes, and 92.8% of these genes (23,971 genes) were functionally annotated. The repeat and gene annotations of the genome assembly and comparative genome analysis between *H. symbiolongicarpus* and *H. vulgaris* highlight a unique genome evolution of *H. symbiolongicarpus*. Our novel chromosome-level genome assembly of *H. symbiolongicarpus* will be a significant resource for the biology of this organism. It also can be useful in comparative genomic studies in stem cell evolution in multicellular organisms.

Data availability

The HSymV2.0 assembly, the gene annotation, and related files have been deposited at Figshare (<https://doi.org/10.6084/m9.figshare.22126232.v1>). The HSymV2.0 assembly has been also deposited at the NCBI under the accession JARBIS000000000. The PacBio HiFi reads are available at the Sequence Read Archive (SRA) under Accession Number SRR23493933. Hi-C reads have been deposited in the SRA with accession SRR23493932.

[Supplemental material](#) available at G3 online.

Acknowledgements

The authors thank Darrin Schultz (University of Vienna, Vienna, Austria), Thea Rogers (University of Vienna, Vienna, Austria), Nina Znidaric (University of Vienna, Vienna, Austria), and Angela Caballero Alfonso (University of Vienna, Vienna, Austria) for their advice on Hi-C experiment. The authors thank Atsushi Ogura (Nagahama Institute of Bio-Science and Technology, Nagahama, Japan), Hiromasa Tabata (Nagahama Institute of Bio-Science and Technology, Nagahama, Japan), and Ikuyo Takemura (Nagahama Institute of Bio-Science and Technology, Nagahama, Japan) for their advice on Hi-C scaffolding analysis. The authors also thank Juan Daniel Montenegro Cabrera (University of Vienna, Vienna, Austria) for his advice regarding genome annotation. We also thank Laura Ryan for animal care, Patricia Calcagno for the image featured in [Fig. 1a](#), and Frank Lab members for the assistance in sample collection. Most computations were performed on the Life Science Compute Cluster (LiSC) of the University of Vienna.

Funding

This work was supported by the Wellcome Trust Investigator award no. 210722/Z/18/Z to U.F., by the Science Foundation Ireland Centre for Research Training in Genomic Data Science (grant no. 18/CRT/6214) to H.R.H., the Austrian Science Fund FWF (I4353) to O.S. and K.K.-N., the Takeda Science Foundation to T.K., the Mochida Memorial Foundation for Medical and Pharmaceutical Research to T.K., and the International Medical Research Foundation to T.K.

Conflicts of interest statement

The authors declare that they have no conflicts of interest.

Literature cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–410. doi:10.1016/S0022-2836(05)80360-2.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.* 2013;69(2):313–319. doi:10.1016/j.ympev.2012.08.023.

Bode HR. The interstitial cell lineage of hydra: a stem cell system that arose early in evolution. *J Cell Sci.* 1996;109(6):1155–1164. doi:10.1242/jcs.109.6.1155.

Bosch TCG, David CN. Stem cells of *Hydra magnipapillata* can differentiate into somatic cells and germ line cells. *Dev Biol.* 1987;121(1):182–191. doi:10.1016/0012-1606(87)90151-5.

Bradshaw B, Thompson K, Frank U. Distinct mechanisms underlie oral vs aboral regeneration in the cnidarian *Hydractinia echinata*. *Elife.* 2015;4:e05506. doi:10.7554/eLife.05506.

Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinform.* 2021;3(1):lqaa108. doi:10.1093/nargab/lqaa108.

Buss LW, Yund PO. A sibling species group of *Hydractinia* in the North-Eastern United States. *J Mar Biol Assoc United Kingdom.* 1989;69(4):857–874. doi:10.1017/S0025315400032215.

Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ.* 2018;6:e4958. doi:10.7717/peerj.4958.

Cazet JF, Siebert S, Little HM, Bertemes P, Primack AS, Ladurner P, AchRAINER M, Fredriksen MT, Moreland RT, Singh S, et al. A chromosome-scale epigenetic map of the *Hydra* genome reveals conserved regulators of cell state. *Genome Res.* 2023;33(2):283–298. doi:10.1101/gr.277040.122.

Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, Rattei T, Balasubramanian PG, Borman J, Busam D, et al. The dynamic genome of *Hydra*. *Nature.* 2010;464(7288):592–596. doi:10.1038/nature08830.

Chen R, Sanders SM, Ma Z, Paschall J, Chang ES, Riscoe BM, Schnitzler CE, Baxevanis AD, Nicotra ML. XY sex determination in a cnidarian. *BMC Biol.* 2023;21(1):32. doi:10.1186/s12915-023-01532-2.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 2021;18(2):170–175. doi:10.1038/s41592-020-01056-5.

Chrysostomou E, Febrimarsa, DuBuc T, Frank U. Gene manipulation in *Hydractinia*. *Methods Mol Biol.* 2022;2450:419–436. doi:10.1007/978-1-0716-2172-1_22.

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. Nanopack: visualizing and processing long-read sequencing data. *Bioinformatics.* 2018;34(15):2666–2669. doi:10.1093/bioinformatics/bty149.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10(2):giab008. doi:10.1093/gigascience/giab008.

Davis MW, Jorgensen EM. Ape, A plasmid editor: a freely available DNA manipulation and visualization program. *Front Bioinform.* 2022;2:818619. doi:10.3389/fbinf.2022.818619.

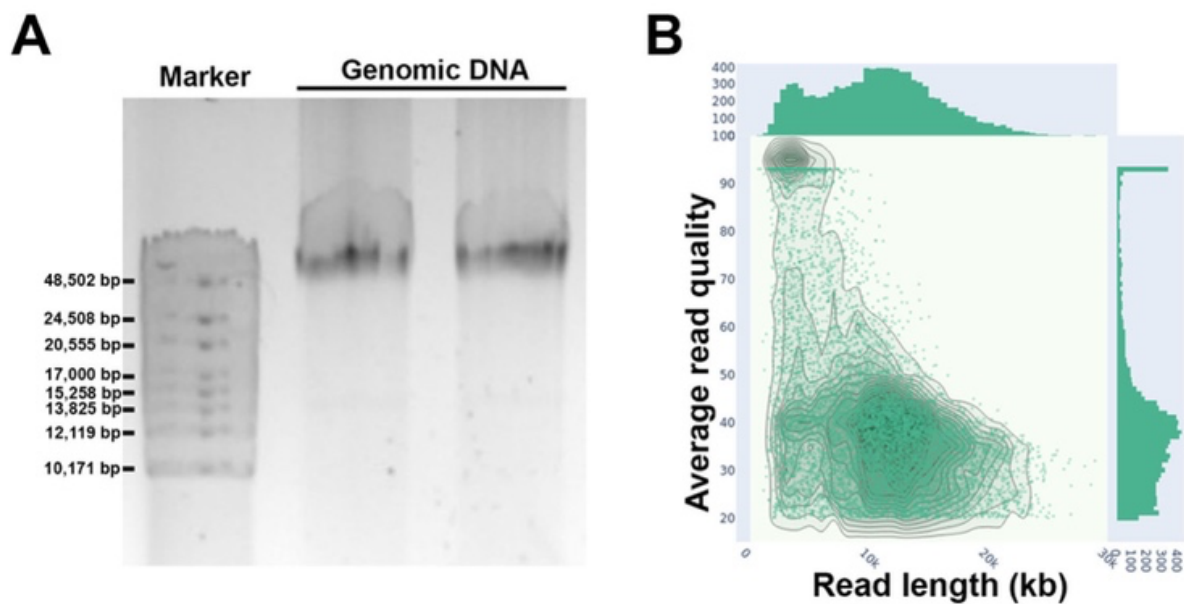
Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21. doi:10.1093/bioinformatics/bts635.

DuBuc TQ, Schnitzler CE, Chrysostomou E, McMahon ET, Febrimarsa GJ, Buggie T, Gornik SG, Hanley S, Barreira SN, Gonzalez P, et al. Transcription factor AP2 controls cnidarian germ cell induction. *Science.* 2020;367(6479):757–762. doi:10.1126/science.aay6782.

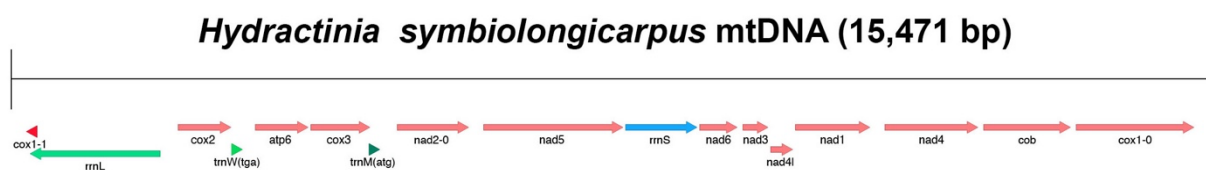
Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. De novo

- assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356(6333):92–95. doi:10.1126/science.aal3327.
- Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 2016a;3(1):99–101. doi:10.1016/j.cels.2015.07.012.
- Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016b;3(1):95–98. doi:10.1016/j.cels.2016.07.002.
- 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 U S A*. 2020;117(17):9451–9457. doi:10.1073/pnas.1921046117.
- Frank U, Nicotra ML, Schnitzler CE. The colonial cnidarian *Hydractinia*. *Evodevo*. 2020;11(1):7. doi:10.1186/s13227-020-00151-0.
- Gabriel L, Hoff KJ, Bruna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*. 2021;22(1):566. doi:10.1186/s12859-021-04482-0.
- Gahan JM, Bradshaw B, Flici H, Frank U. The interstitial stem cells in *Hydractinia* and their role in regeneration. *Curr Opin Genet Dev*. 2016;40:65–73. doi:10.1016/j.gde.2016.06.006.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072–1075. doi:10.1093/bioinformatics/btt086.
- Hobmayer B, Jenewein M, Eder D, Eder MK, Glasauer S, Gufler S, Hartl M, Salvenmoser W. Stemness in *Hydra*—a current perspective. *Int J Dev Biol*. 2012;56(6–7–8):509–517. doi:10.1387/ijdb.113426bh.
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(D1):D309–D314. doi:10.1093/nar/gky1085.
- Juliano CE, Reich A, Liu N, Götzfried J, Zhong M, Uman S, Reenan RA, Wessel GM, Steele RE, Lin H. PIWI Proteins and PIWI-interacting RNAs function in *Hydra* somatic stem cells. *Proc Natl Acad Sci U S A*. 2014;111(1):337–342. doi:10.1073/pnas.1320965111.
- Künzel T, Heiermann R, Frank U, Müller W, Tilmann W, Bause M, Nonn A, Helling M, Schwarz RS, Plickert G. Migration and differentiation potential of stem cells in the cnidarian *Hydractinia* analysed in eGFP-transgenic animals and chimeras. *Dev Biol*. 2010;348(1):120–129. doi:10.1016/j.ydbio.2010.08.017.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094–3100. doi:10.1093/bioinformatics/bty191.
- Lin J, Feng S, Wang L. Complete mitochondrial genome sequence of *Aequorea coerulescens*. *Mitochondrial DNA Part B Resour*. 2019;4(2):2882–2883. doi:10.1080/23802359.2019.1660257.
- Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. *F1000Res*. 2020;9:ISCB Comm J-304. doi:10.12688/f1000research.23297.
- Plickert G, Frank U, Müller WA. *Hydractinia*, a pioneering model for stem cell biology and reprogramming somatic cells to pluripotency. *Int J Dev Biol*. 2012;56(6–7–8):519–534. doi:10.1387/ijdb.123502gp.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–842. doi:10.1093/bioinformatics/btq033.
- Quiroga-Artigas G, Duscher A, Lundquist K, Waletzki J, Schnitzler CE. Gene knockdown via electroporation of short hairpin RNAs in embryos of the marine hydroid *Hydractinia symbiolongicarpus*. *Sci Rep*. 2020;10(1):12806. doi:10.1038/s41598-020-69489-8.
- Rice P, Longden I, Bleasby A. EMBOS: the European Molecular Biology Open Software Suite. *Trends Genet*. 2000;16(6):276–277. doi:10.1016/S0168-9525(00)00204-2.
- Sanders SM, Ma Z, Hughes JM, Riscoe BM, Gibson GA, Watson AM, Flici H, Frank U, Schnitzler CE, Baxevanis AD, et al. CRISPR/Cas9-mediated gene knockin in the hydroid *Hydractinia symbiolongicarpus*. *BMC Genomics*. 2018;19(1):649. doi:10.1186/s12864-018-5032-z.
- Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2020;37(12):1639–1643. doi:10.1093/bioinformatics/btaa1016.
- Simakov O, Bredeson J, Berkoff K, Marletaz F, Mitros T, Schultz DT, O'Connell BL, Dear P, Martinez DE, Steele RE, et al. Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci Adv*. 2022;8(5):eabi5884. doi:10.1126/sciadv.abi5884.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–3212. doi:10.1093/bioinformatics/btv351.
- 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*. 2023;33(10):1883–1892.e3. doi:10.1016/j.cub.2023.03.039. Epub 2023 Apr 6.
- Weismann A. The Origin of the Sexual Cells in Hydromedusae (Foreign Title: Die Entstehung der Sexualzellen bei Hydromedusen). Jena: Gustav Fischer; 1883.
- Zdobnov EM, Apweiler R. Interproscan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 2001;17(9):847–848. doi:10.1093/bioinformatics/17.9.847.
- Zou H, Zhang J, Li W, Wu S, Wang G. Mitochondrial genome of the freshwater jellyfish *Craspedacusta sowerbyi* and phylogenetics of Medusozoa. *PLoS One*. 2012;7(12):e51465. doi:10.1371/journal.pone.0051465.

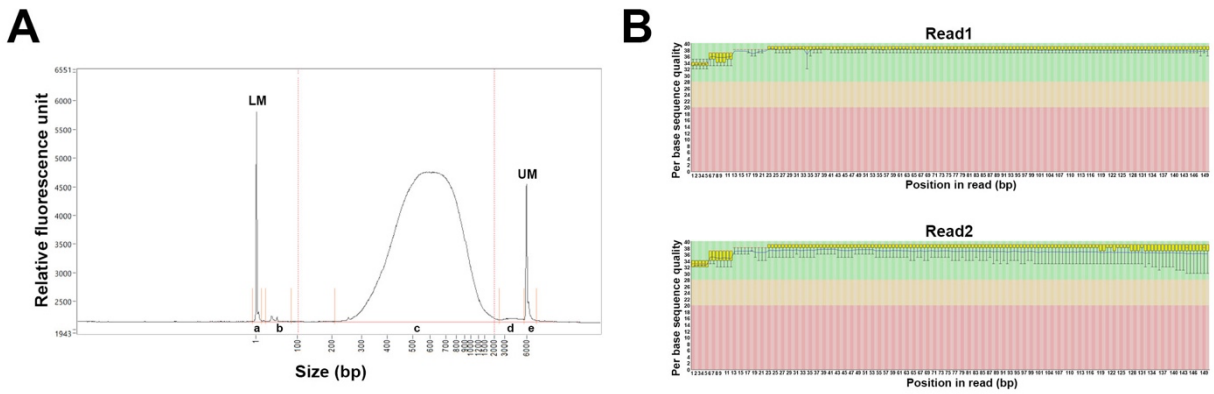
Editor: J. Parker



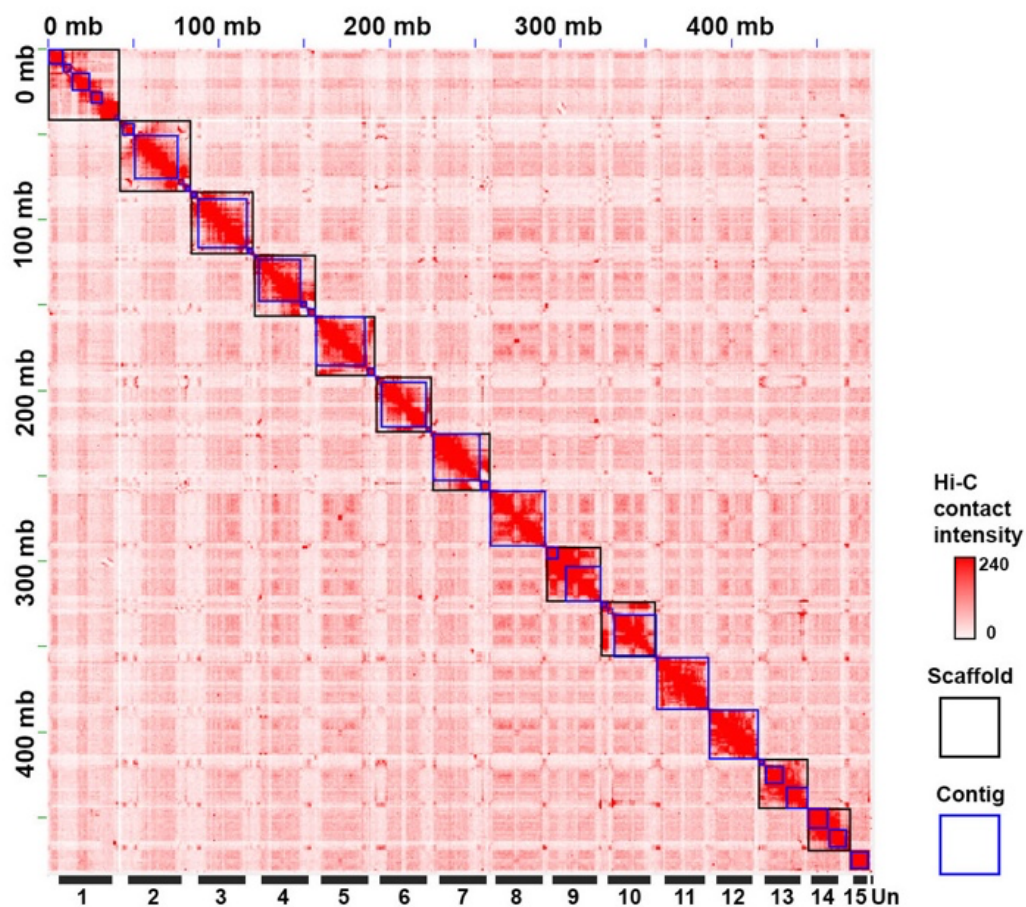
Supplementary Figure 1. PacBio HiFi long-read sequencing of the *H. symbiolongicarpus* genome. (A) Agarose gel electrophoresis of the extracted genomic DNA. (B) Distribution of read lengths and average read qualities of the PacBio HiFi long-reads. The solid lines are kernel density estimation.



Supplementary Figure 2. Mitochondrial DNA of *H. symbiolongicarpus*. Arrows indicate gene orientations.



Supplementary Figure 3. Hi-C sequencing of the *H. symbiolongicarpus* genome. (A) Insert size distribution. The insert size distribution is approximately between 350 bp and 1,000 bp. The DNA concentrations of the DNA fragments in a (from 0 bp to 15 bp), b (from 25 bp to 84 bp), c (from 210 bp to 2,474 bp), d (from 2,474 bp to 5,698 bp), e (5,698 bp to 7,453 bp) are 0.0125 (ng/ μ L), 0.0198 (ng/ μ L), 4.27 (ng/ μ L), 0.0240 (ng/ μ L), 0.0058 (ng/ μ L), respectively. LM at 1 bp denotes the lower marker and UM at 6,000 bp denotes the upper marker. (B) Per base sequence quality of the Hi-C sequencing reads.



Supplementary Figure 4. Hi-C contact map of the HSymV2.0 assembly with the boundaries of scaffolds (black squares) and contigs (blue squares). Note that scaffold 8, 11, 12, and 15 consist of single contigs.

Results

Paper III

Title: “Subtelomeric repeat expansion in Hydractinia symbiolongicarpus chromosomes”

Authors:

Tetsuo Kon^{1*#}, Koto Kon-Nanjo^{1#}, Oleg Simakov¹

¹ Department of Neurosciences and Developmental Biology, University of Vienna, Vienna, 1030, Austria

*Correspondence to Tetsuo Kon (kontetsuo@gmail.com)

#Equally contributed

Status:

Published in Mobile DNA, 2025, 16(1):14. doi: 10.1186/s13100-025-00355-y.
doi: 10.1093/g3journal/jkad107.

BRIEF REPORT

Open Access



Subtelomeric repeat expansion in *Hydractinia symbiolongicarpus* chromosomes

Tetsuo Kon^{1*}, Koto Kon-Nanjo^{1†} and Oleg Simakov¹

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

¹Department of Neurosciences and Developmental Biology, University of Vienna, Vienna 1030, Austria



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

[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 rmbast 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 HSymSHIR (*H. symbiolongicarpus* subtelomeric *Helitron*-like repeat). The copy number of the HSymSHIR 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 HSymIS3EUIR (*H. symbiolongicarpus* IS3EU-like repeat). The copy number of the HSymIS3EUIR was 1,384 (Table S3). Moreover, the Kimura substitution level of HSymSHIR 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 HSymIS3EUIR showed a Kimura substitution level of 20.6%, which corresponds to the second wide peak (Fig. 1b, Table S3). In summary, we identified HSymSHIR, HSymSR, and HSymIS3EUIR

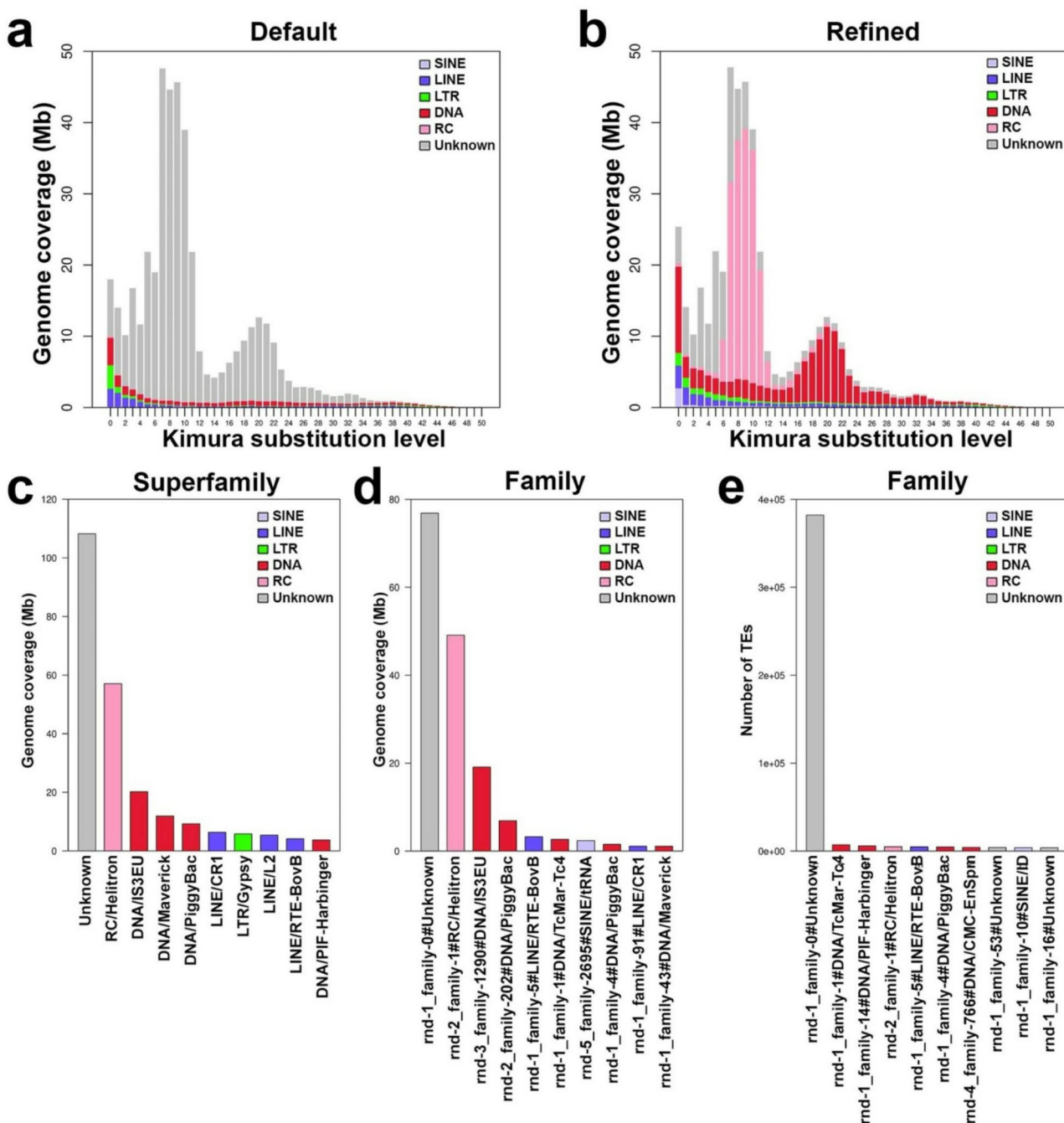


Fig. 1 TE genome coverage in *H. symbiolongicarpus* genome. **(a)** Repeat landscape illustrating TE accumulation history for the *H. symbiolongicarpus* genome with the custom repetitive element library generated by RepeatModeler run with the default settings. The horizontal axis illustrates the Kimura 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.*

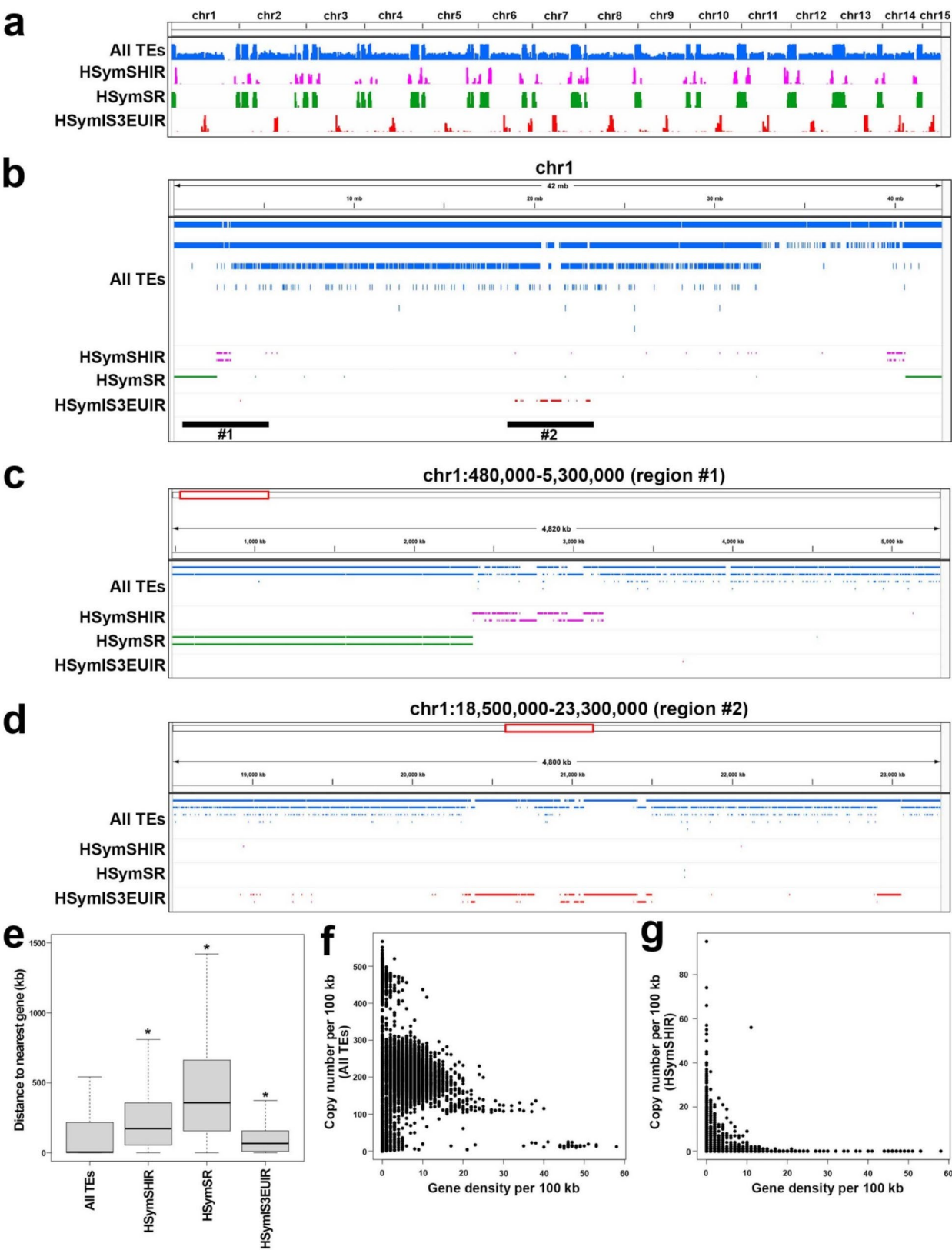


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Subtelomeric and centromeric 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 $< 2.2 \times 10^{-16}$ 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 HSymIS3EUIR 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 HSymSHIR 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 HSymSHIR 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 HSymSHIR 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 HSymSHIR (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

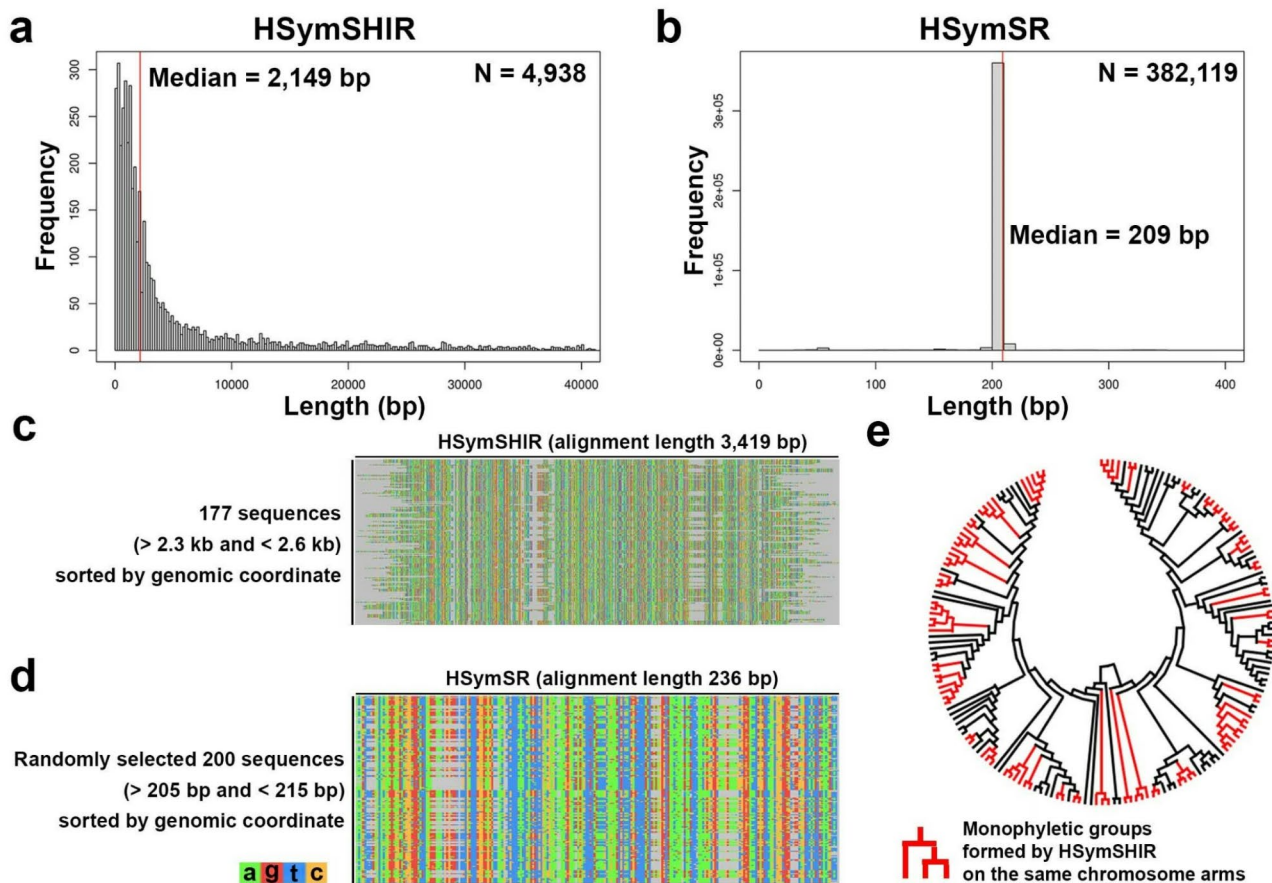


Fig. 3 Molecular phylogenetic analysis of HSymSHIR in *H. symbiolongicarpus* genome. **(a)** Sequence length distribution of HSymSHIR. **(b)** Sequence length distribution of HSymSR. **(c)** Multiple sequence alignment of all 177 HSymSHIR, with sequence lengths greater than 2.3 kb and less than 2.6 kb. **(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 HSymSHIR sequences from the same chromosome arms are highlighted in red

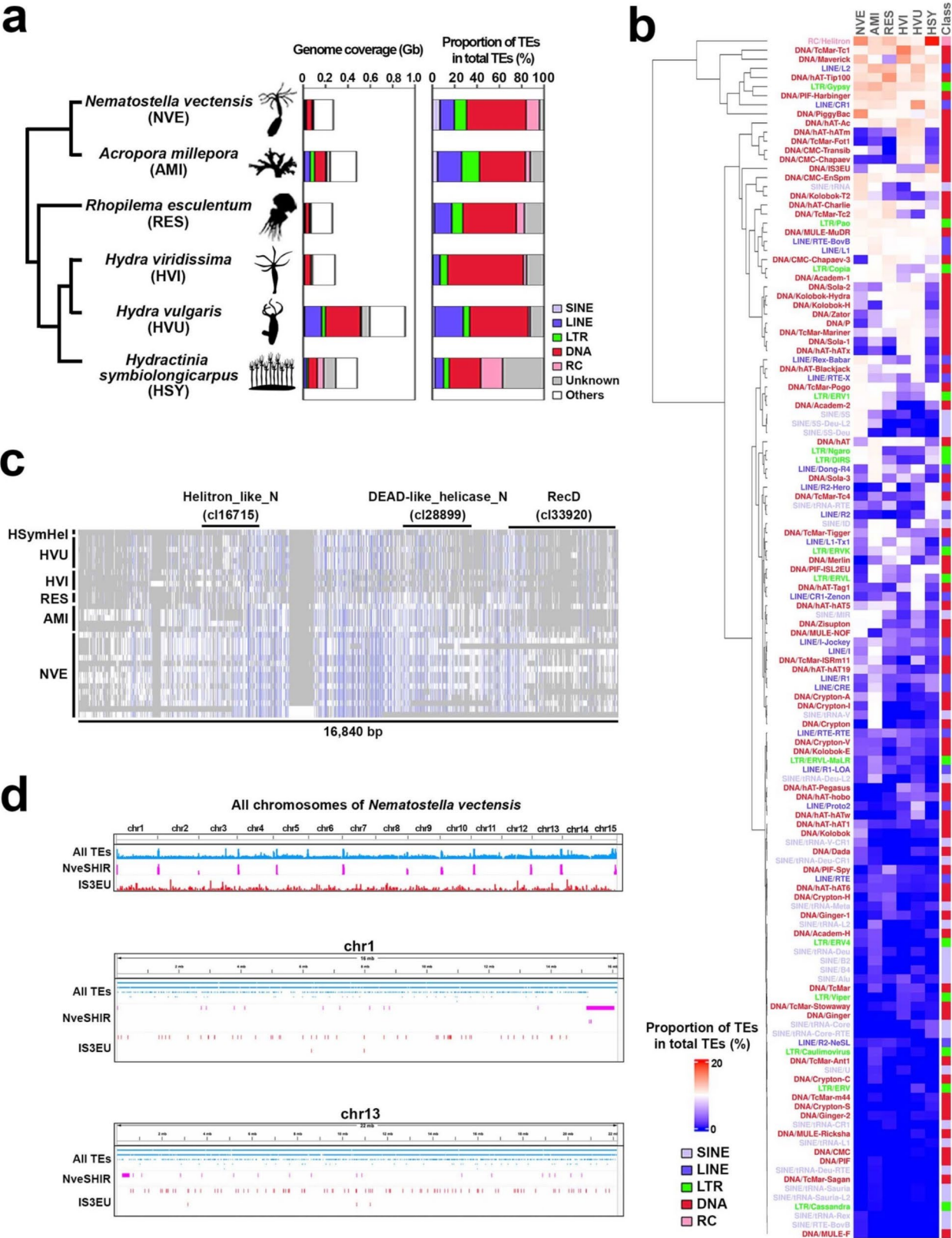


Fig. 4 (See legend on next page.)

(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 NveSHIR in subtelomeric regions. The first track shows the distribution of all TEs, the second track shows the distribution of NveSHIR, 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 (NveSHIR) 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 HSymSHIR, HSymSR, and HSymIS3EUIR, 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 HSymSHIR

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	<i>Acropora millepora</i>
HEL1	<i>Helitron</i> -like element 1
HEL2	<i>Helitron</i> -like element 2
HMM	Hidden Markov Model
HSY	<i>Hydractinia symbiolongicarpus</i>
HSymHel	<i>H. symbiolongicarpus</i> Helitron
HSymSHIR	<i>H. symbiolongicarpus</i> subtelomeric <i>Helitron</i> -like repeat
HSymIS3EUIR	<i>H. symbiolongicarpus</i> IS3EU-like repeat
HSymSR	<i>H. symbiolongicarpus</i> subtelomeric repeat
HVU	<i>Hydra vulgaris</i>
HVI	<i>Hydra viridissima</i>
NveSHIR	<i>N. vectensis</i> subtelomeric <i>Helitron</i> -like repeat
NVE	<i>Nematostella vectensis</i>
RC	Rolling-Circle
RCR	Rolling-Circle Replication
RES	<i>Rhopilema esculentum</i>
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. Scharf 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 A, 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, Khalitun 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 Conceição 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, Martellosi 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, Bilal 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. Repeat-Modeler2 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 (Helitrons). *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/dnares/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://mobilejournal.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, McPeck 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, Propf 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://mobilejournal.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

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary figures

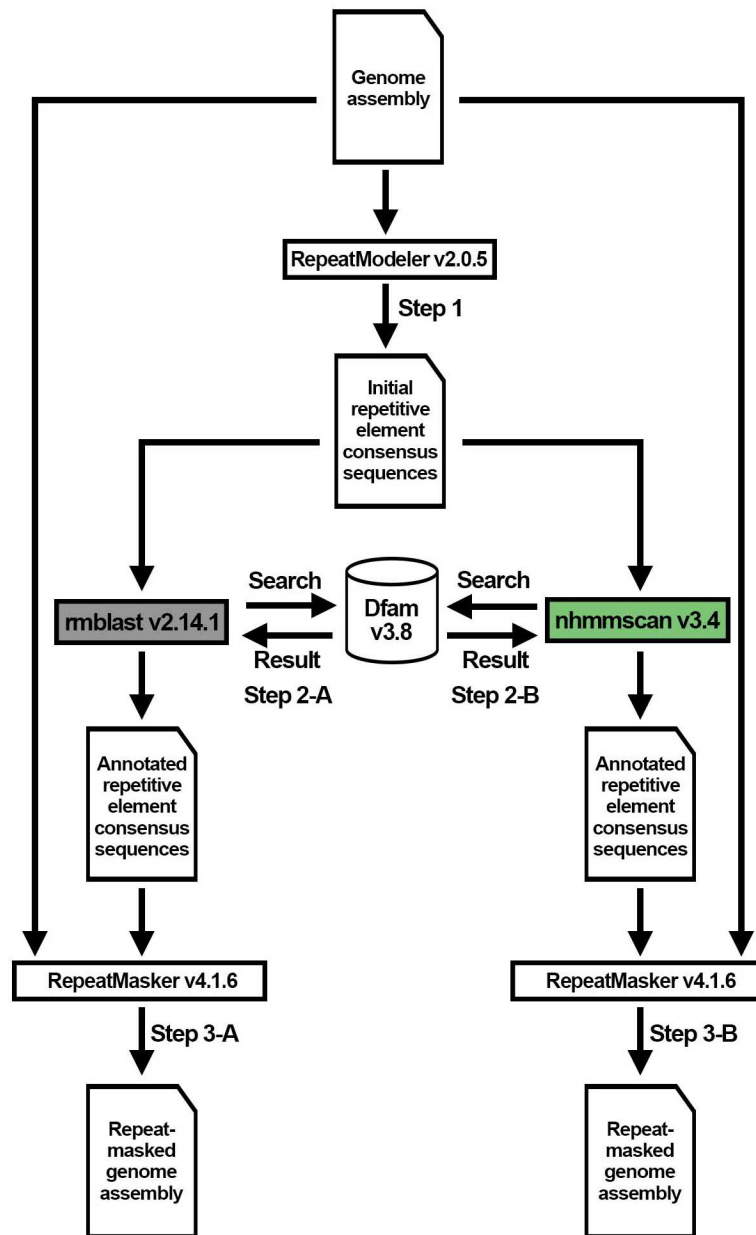


Fig. S1. Pipeline for analyzing repetitive elements in genome assemblies

The pipeline for repetitive element analysis is presented as a flowchart. Step 1 represents the common initial step. Steps 2-A and 3-A correspond to the default RepeatModeler/RepeatMasker workflow, whereas Steps 2-B and 3-B indicate the modified steps implemented in this study.

	DR003132626	100	CTTAACCGATTCTAACCTTTACAGGTCTCAATCGATTCTGTCCTGCACGGGAGTGTGTAGCTATGTAGNA. TATCGAAAAA	181
rnd-2_family-1#RC/Helitron		779	+T+AA TAC CC+ AC TCAATCG TTC C GT+ A AG G TG A C TAG+ A TATCG+AAAAA	
		779	TTCAATTAAGTCCACAG-----TCAATCGTTTTCA--CTGTTA--AAAGCGATGCAAC--CTAGTAATATCGGAAAAA	713
			5555555555555555443.....678888777777...445554...6788888888754...4566555****9887655	PP
	DR003132626	182	AACAAAAATGACGGGAGATACAGGNTGNAAAAAACGTTTACGTGCTTAATCG.....CTGAATT	240
rnd-2_family-1#RC/Helitron		712	-----TATCGCTAAATACATGTTGGTACGAA----TATGTGACGCAACCGTgtgaactcaaacggattgacaccG-----T	646
			56688889999*****99888888....5788888888888822222222222233311111110.....2	PP
	DR003132626	241	CAGGAAGGGAATG...CGTCTATTTT.....CTTTTATTACAG..GCATTTACCTGTAAGGTTTTCGACGGCGTCA	307
rnd-2_family-1#RC/Helitron		645	AG AA AATG +G TA TTT TCT TT +T CAGC GC++T AC+ TAGG TT C +CG A	
		645	TAGAAANTCAATGaaTG--TAGTTTttgaaactgtcTCTATTCTTCAGCgcGCGCTAACGGCTAGGATTGGTTCACGGTA	565
			33434333333333233..4433331122222212456677778887611567899*****97778888999*	PP
	DR003132626	308	AAAANGATTAAATAAAATAATATTNAAAGAAATGATACCCGA	351
rnd-2_family-1#RC/Helitron		564	AAAA GA T A T A TAAAT TT+AA AAA G TACC A	
		564	AAAATGACTTATTTATCTAAATTTTTAANTAAGCGCTACCA	521
			*****9887 PP	

DR001866860 154 CGGCGTTATAACGCCCTAA...TAC.....GGCGTTATAACGCCCTAATACGCCGTTATAACGCCCTAATTTAGGGCGT 224
 +GGC TTATAAC CC T A TAC GGC T AA C +TAATA GGCCT T GCCC TAA T A GGC T
 rnd-3_family-1290#DNA/IS3EU 1685 TGGCCTTATAACACCGCTCAacTACgtgcacAGCCCTCTAAAGNCGTGAATATGGCGTTTGGCCGCC--TAA--TAAAGCGCT 1762
 2333333333333333322112221111113334344444444555555555555555555555554.333.344455555 PP

DR001866860 225 TATAACGCCGTAATTACGGCGTTATAACGCCGTAATAGGGCGTTAT.AACGCCGTAA.....ATTAG 285
 TA A C AAT AC GCG T ACG + G GTTA A CG TAA A TA
 rnd-3_family-1290#DNA/IS3EU 1763 TACA---CGTAATACAC--GGCATTCACGTGCCCGNCGAG--GTTAACACGGCTCAATctctccagttctcaatgagACTAC 1837
 5322...2222222233.2444444444322221122222.2222213333333333444444444444444444466677 PP

DR001866860 286 GGC GTTATAACGCCGTAAATT.....AGGGCGTTATAACGCCGTAAATTA.....GGGCGTTATAACGCCGT 347
 G CGT AA CGT T A G C TTA AAC CCGT AA TA GGC T AACGC+GT
 rnd-3_family-1290#DNA/IS3EU 1838 GACGTCCCAAGT--CGTT---TtccttngcgccANGTCCTTACAACACCGTCAACTAcggcacAGGCCCTCTAAACGCTGT 1913
 777776666543.3332...123334444446677777777777777777777777552222234666666677777777 PP

DR001866860 348 A.TAGGGCGTTATAACGCCGTATTAGGGCGTNATAACGCCGTA 389
 A TA GGC TT GCC AT AGGGC T+A A CGTA
 rnd-3_family-1290#DNA/IS3EU 1914 AaTATGGCCTTTNGNAGCCATAAAGGGCCTTACA----CGTA 1952
 436666666666666677667777777777666654...4444 PP

(a) Consensus of the HSymSHIR (rnd-2_family-1) and the *Helitron* sequence from Dfam (accession number DR003132626). **(b)** Consensus sequence of HSymIS3EUIR (rnd-3_family-1290) and the corresponding IS3EU sequence from Dfam (accession number DR001866860).

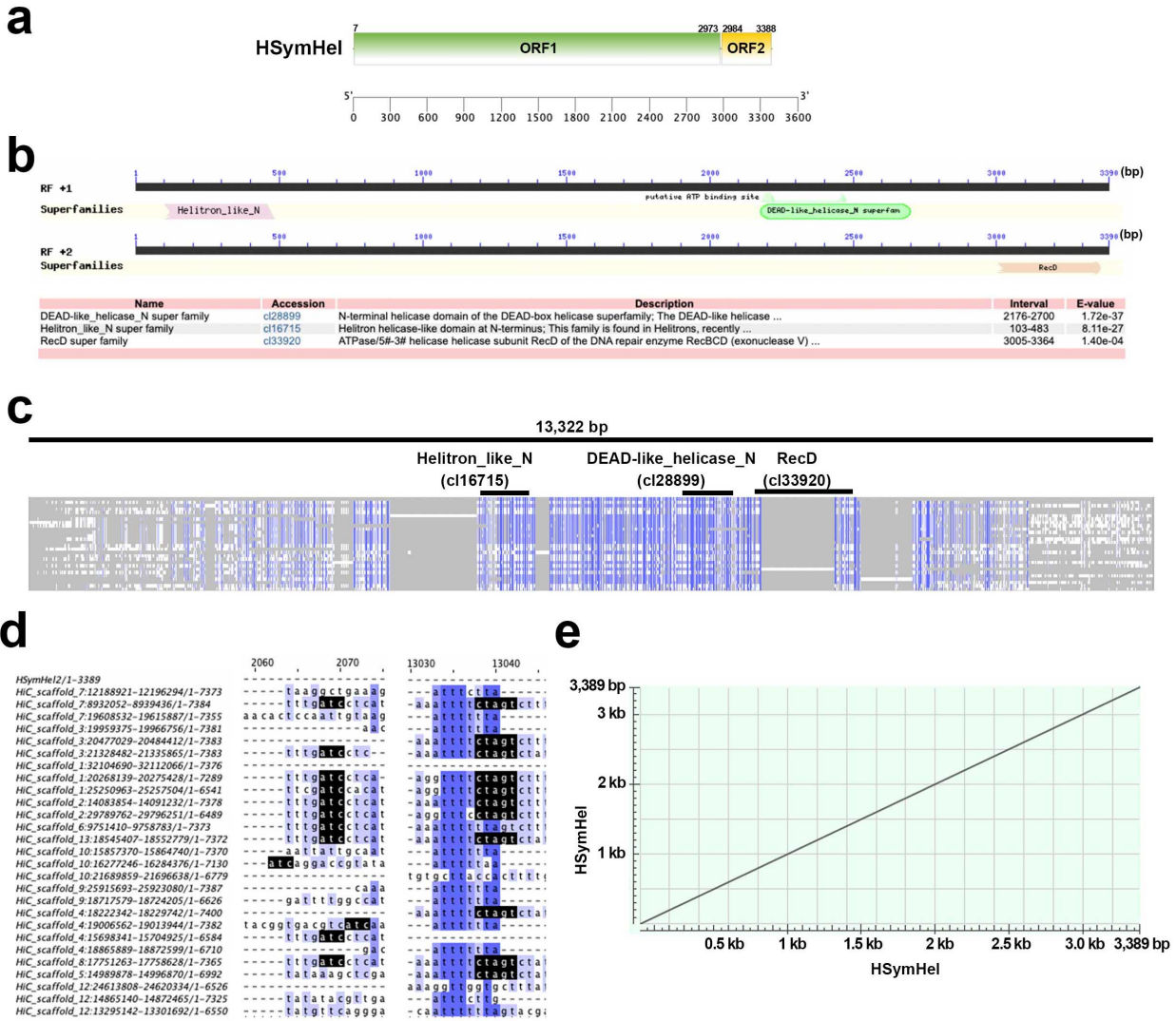


Fig. S3. Characterization of HSymHel

(a) Open reading frames (ORFs) in HSymHel. HSymHel contains two ORFs (ORF1 and ORF2). (b) Domain organization of HSymHel. Screenshot of the result from the NCBI Conserved Domain Search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) are shown. (c) Multiple sequence alignment. Alignment of the HSymHel consensus sequence and its copies in the genome. The domain names are determined based on the results from the NCBI Conserved Domain Search analysis. (d) Alignment of putative genomic insertion sites of HSymHel. The potential transition sites between the genome and the HSymHel sequence are indicated in white. (e) Dot plot showing a self-sequence alignment of the HSymHel consensus sequence.

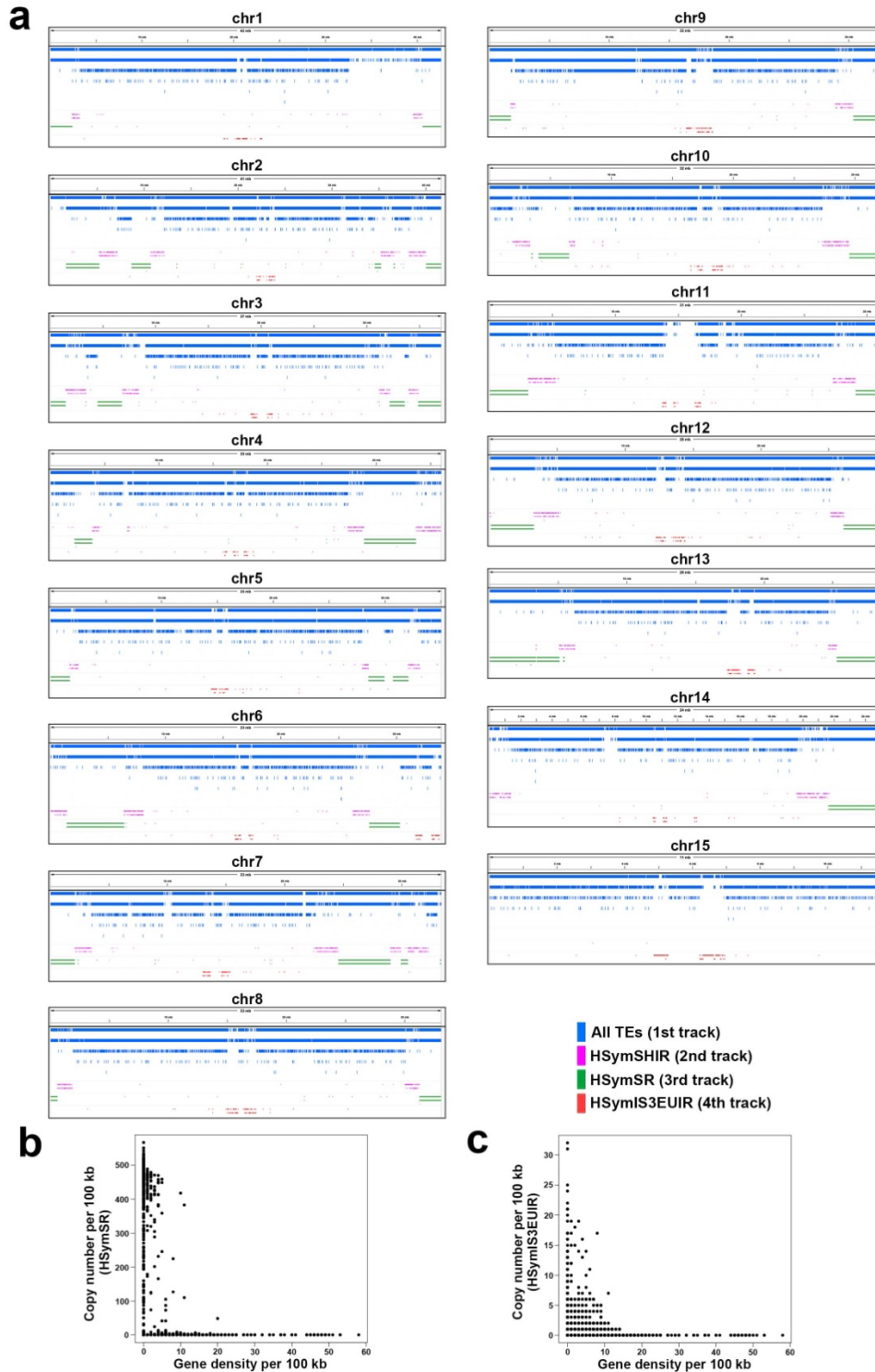


Fig. S4. Selected repeat localization in each chromosome

(a) The first track represents all TEs, the second track represents HSymSHIR, and the third track depicts HSymSR. The fourth track shows HSymIS3EUR. (b) The relationship between gene density and HSymSR density. (c) The relationship between gene density and HSymIS3EUR density.

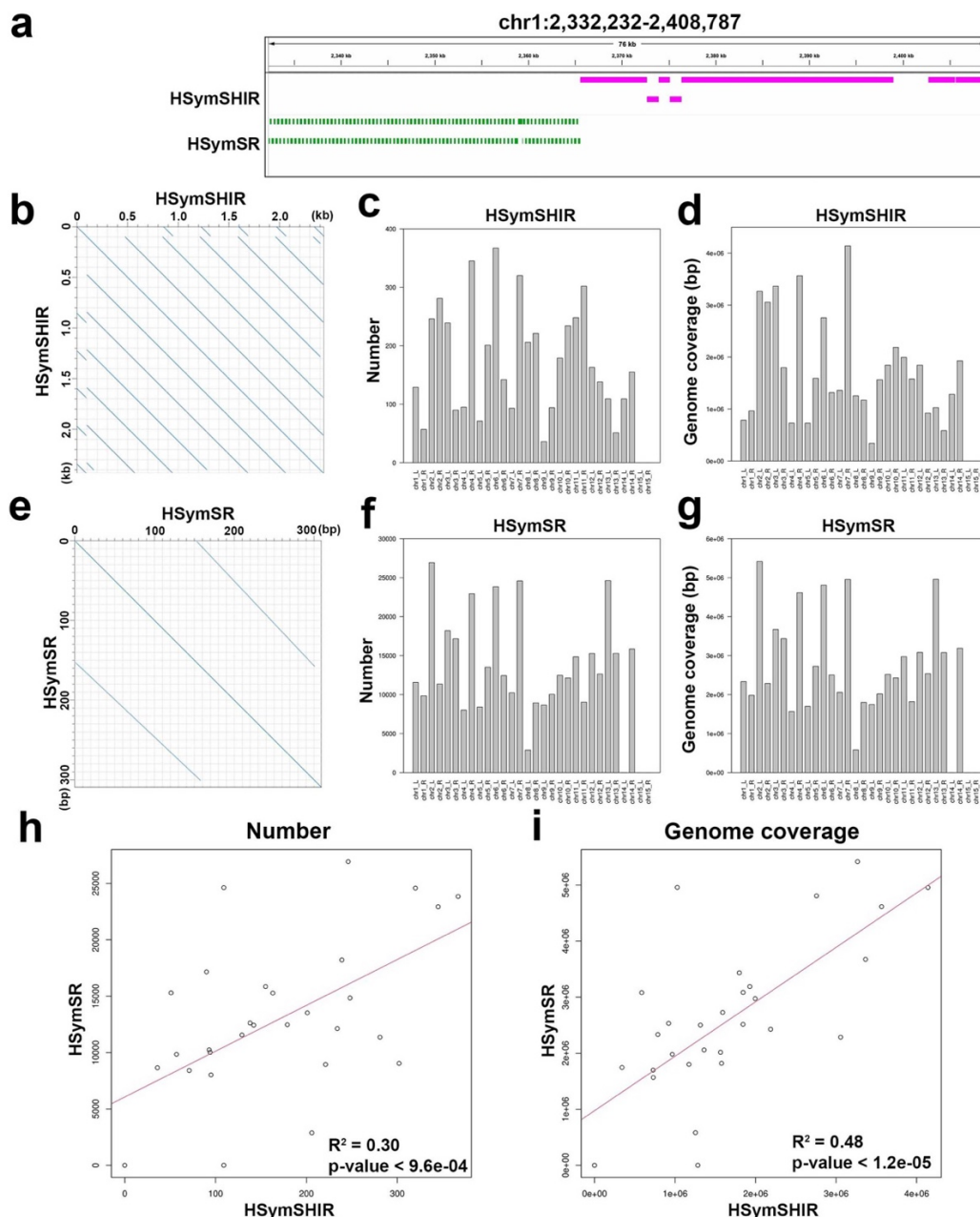


Fig. S5. Copy numbers and genome coverage of the single *Helitron* family

(a) Regional distribution of HSymSHIR and HSymSR on chromosome 1 base from 2,326,326 bp to 2,402,881 bp. (b) Self-sequence alignment of the consensus sequence of HSymSHIR. The consensus sequence of HSymSHIR contains a repetitive sequence with a unit length of 372 bp. (c) Number of HSymSHIR in each chromosome arm. (d) Genome coverage of HSymSR in each chromosome arm. (e) Self-sequence alignment of the consensus sequence of HSymSR. The consensus sequence of HSymSR contains a repetitive sequence with a unit length of approximately 150 bp. (f) Number of HSymSR in each chromosome arm. (g) Genome coverage of HSymSR in each chromosome arm. (h) Correlation analysis between number of HSymSHIR and that of HSymSR in each chromosome arm. (i) Correlation analysis between genome coverage of HSymSHIR and that of HSymSR in each chromosome arm.

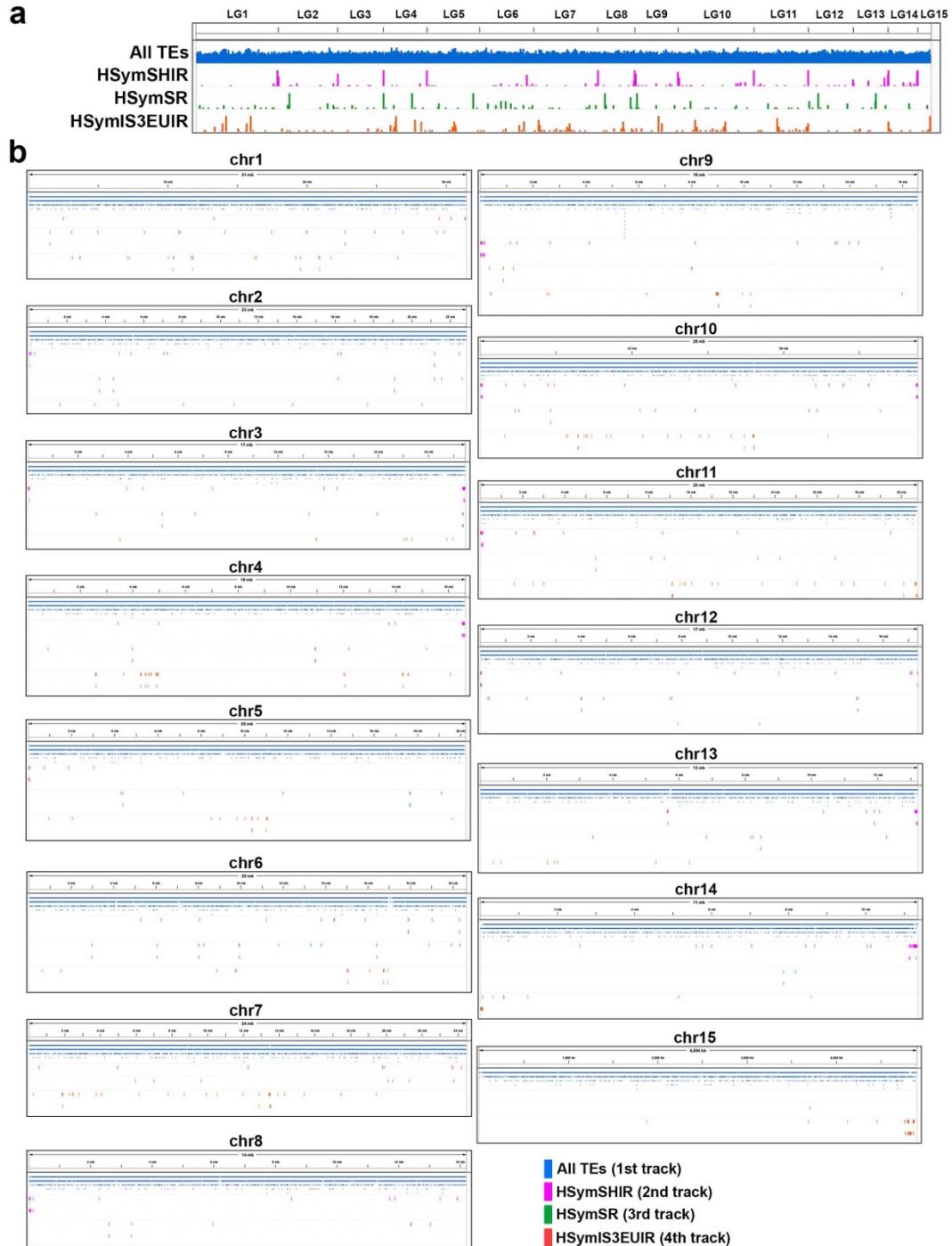


Fig. S6. *Helitron* localization on the pseudochromosomes generated by the Hsym_primary_v1.0 assembly and the maternal genetic linkage map

(a) *Helitron* distribution across all pseudochromosomes. Pseudochromosomes were generated by the Hsym_primary_v1.0 assembly and the maternal genetic linkage map. Note the accumulation of HSymSHIR at the ends of chromosomes.

(b) *Helitron* distribution in each pseudochromosomes. The track arrangement is the same as in panel (a).

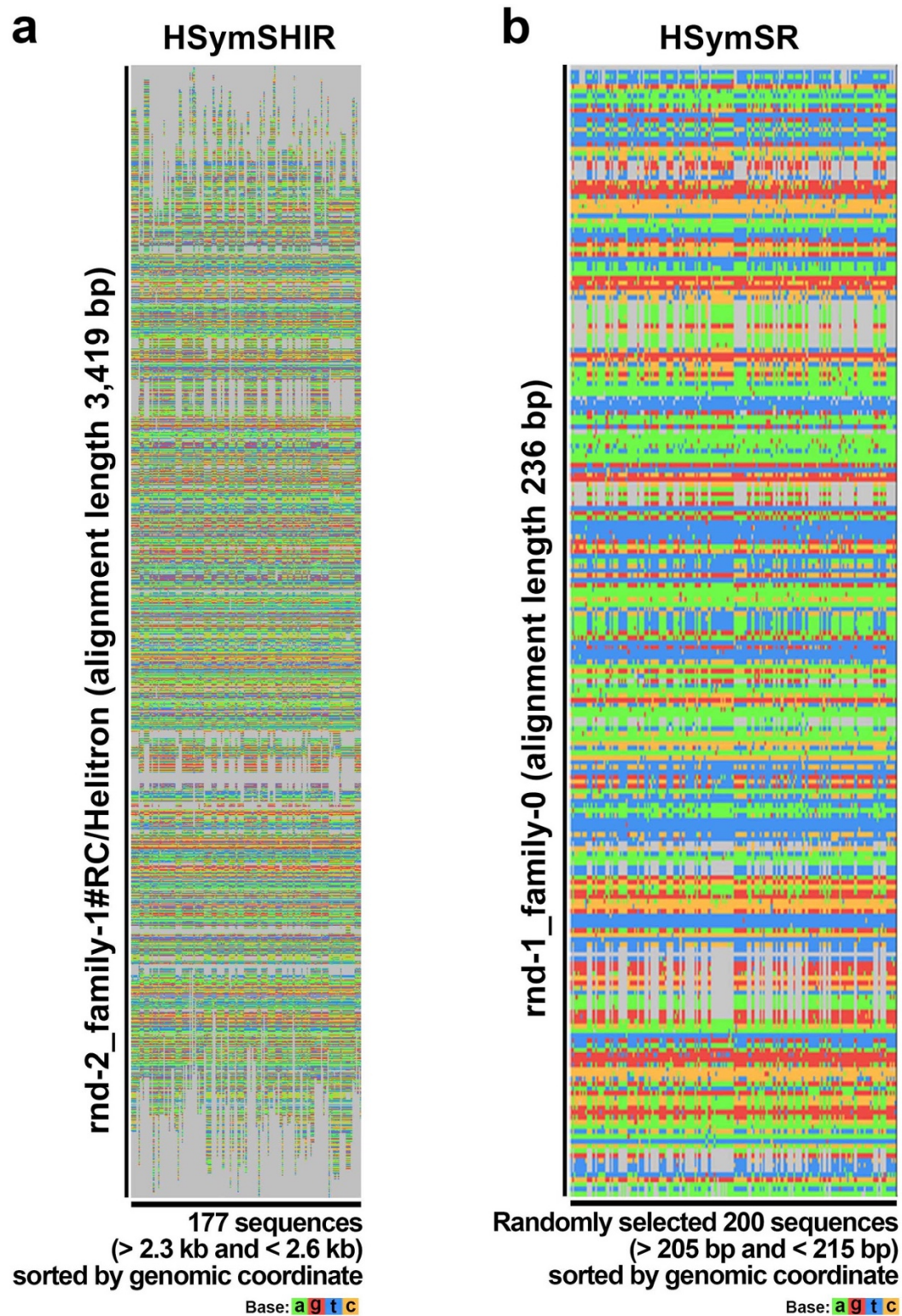


Fig. S7. Multiple sequence alignments of HSymSHIR and HSymSR

Panels (a) and (b) are enlarged views of the sequence alignments in Fig. 3c and Fig. 3d.

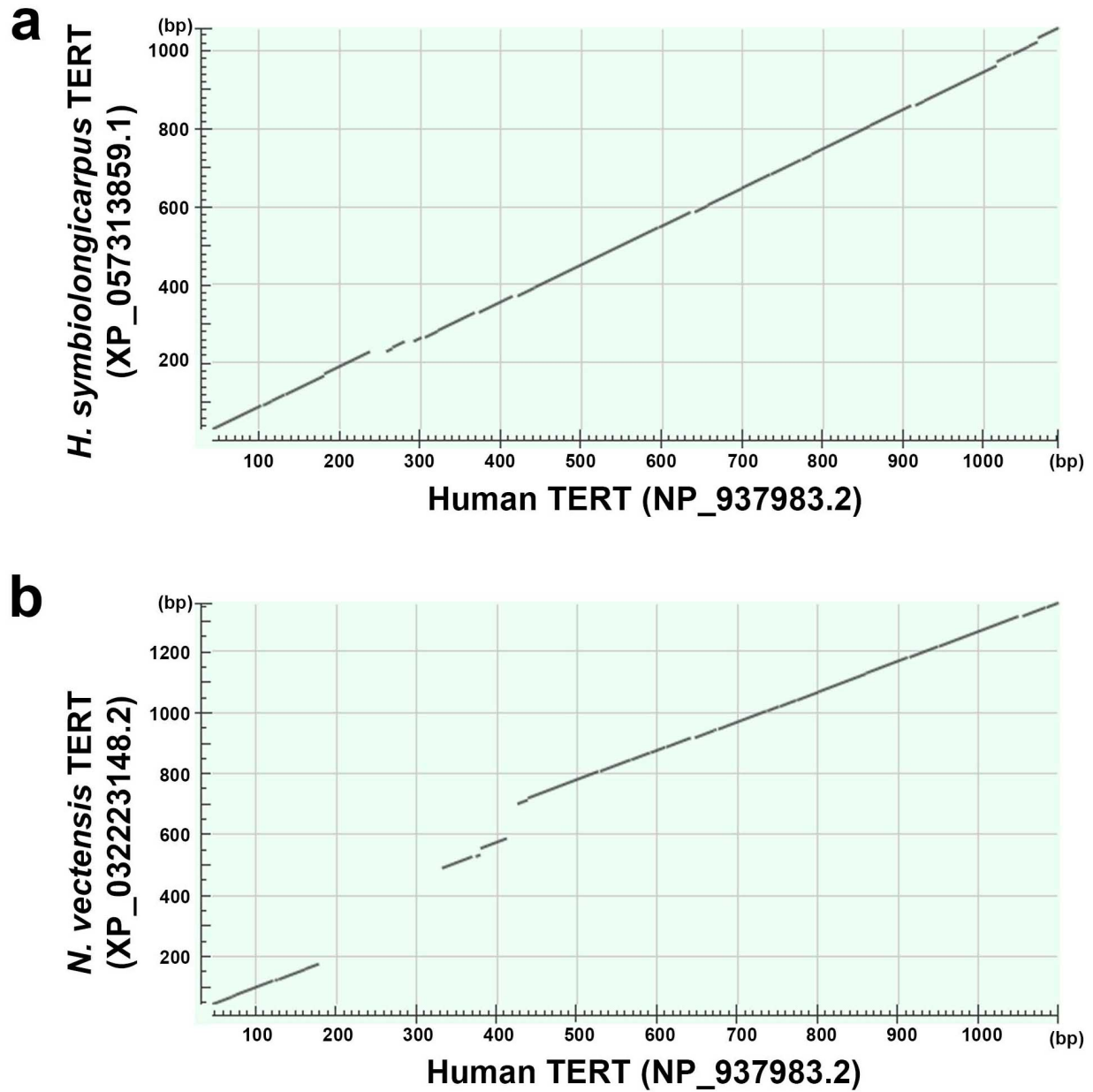


Fig. S8. Dot plots showing the protein sequence alignment between human TERT and the TERT of *H. symbiolongicarpus* or *N. vectensis*

(a) Protein sequence alignment between human TERT (horizontal axis) and *H. symbiolongicarpus* TERT (vertical axis).

(b) Protein sequence alignment between human TERT (horizontal axis) and *N. vectensis* TERT (vertical axis).

Supplementary Tables

Table S1. Genome coverage of major TE groups

Table S2. Genome coverage of TE superfamilies

Table S3. Genome coverage of TE families

Table S4. List of cnidarian species analyzed in this study

Concluding Discussion

Together, the studies presented in this thesis demonstrate that transposable elements (TEs) act as both conserved and dynamic components of animal chromosomes. They contribute to genome diversification while maintaining large-scale chromosomal stability, thereby bridging the long-standing gap between chromosomal conservation and evolutionary plasticity. Through an integrative approach that combines chromosome-scale genomics, stem-cell-resolved analyses, and cross-species comparisons, this work provides new insights into how genome architecture and TE dynamics coevolve across early-diverging metazoans.

The first section of the Result chapter comprises the publication “The dynamic genomes of *Hydra* and the anciently active repeat complement of animal chromosomes” (Kon-Nanjo *et al*, 2025). In this study, I generated haplotype-resolved, telomere-to-telomere genome assemblies of two genetically divergent *Hydra vulgaris* strains and performed genomic and transcriptomic analyses across the organism’s three stem-cell lineages. This approach revealed that distinct TE families are transcriptionally and genomically active in each lineage through non-random insertion patterns. Importantly, a deeply conserved core set of DNA transposons was identified as an anciently active component of animal chromosomes, contributing to gradual genomic expansion across metazoans. These elements influence structural variation around loci associated with cell proliferation and chromatin topology, suggesting that anciently active DNA transposons act as self-renewing components of chromosome architecture. Beyond *Hydra*, these findings provide a conceptual framework for understanding how deeply conserved chromosomal structures can accommodate continuous TE-mediated innovation without compromising overall genome integrity.

The second section of the Result chapter comprises the publication “Chromosome-level genome assembly of *Hydractinia symbiolongicarpus*” (Kon-Nanjo *et al*, 2023). In this study, I produced the first chromosome-level genome assembly of the colonial hydrozoan *Hydractinia symbiolongicarpus* using PacBio HiFi long-read sequencing combined with Hi-C scaffolding. The resulting 483 Mb assembly, organized into 15 chromosomes representing 99.8% of the genome, revealed that approximately 61% consists of repetitive sequences, reflecting at least two distinct phases of repeat expansion. The genome encodes 25,825 protein-coding genes with 93.1% BUSCO completeness and exhibits extensive macrosynteny conservation with *Hydra vulgaris*. Together, these results demonstrate that, despite considerable variation in genome size, hydrozoans maintain large-scale chromosomal stability. The *Hydractinia* genome therefore serves as a valuable resource for exploring how repeat accumulation and gene regulation interact during stem-cell maintenance and differentiation in colonial organisms. This study also laid the foundation for future cross-species comparisons aimed at uncovering the genomic transitions that accompanied the evolution of pluripotent stem-cell systems within cnidarians.

The last section of the Result chapter comprises “Expansion of a single Helitron subfamily in *Hydractinia symbiolongicarpus* suggests a shared mechanism of cnidarian chromosomal extension” (Kon-Nanjo *et al*, 2025). I discovered an extensive subtelomeric accumulation of a single 372 bp repeat unit related to the Helitron TE family in *H. symbiolongicarpus*. Together with a tandem minisatellite repeat, this element family accounts for 26% of the total genome and largely explains its size expansion relative to other cnidarians. Comparative analyses showed that other Helitron-like sequences also accumulate in the subtelomeric regions of the *Nematostella* genome, indicating that localized

amplification of these elements may drive subtelomeric extension and lineage-specific genome expansion within the cnidaria.

Collectively, these findings establish that TE activity operates as a continuous and multifaceted evolutionary force. Rather than being transient genomic invaders, TEs may have functioned as long-term architects of chromosomal structure and regulatory potential. By integrating high-resolution genomic analyses with lineage-specific perspectives, this thesis highlights how the interplay between TE activity, stem-cell maintenance, and chromosomal conservation has shaped genome evolution across animals. The evidence presented here supports the view that genome evolution is not solely governed by the accumulation of mutations but also by the dynamic equilibrium between stability and flexibility maintained through TE-driven remodeling.

In light of these findings, a broader evolutionary perspective emerges in which the gradual “slow-mode” genome expansions mediated by A-TEs, which are mainly composed of DNA elements, may represent a distinct evolutionary regime compared with the rapid, burst-like expansions driven by retroelements. Such DNA element-based mechanisms likely promote incremental chromosomal growth over long timescales, allowing the preservation of regulatory landscapes and the emergence of subtle changes in genome topology. In contrast, retroelement-driven bursts can cause large-scale reorganization and rapid remodeling of enhancer-promoter interactions. These contrasting modes of genome expansion may therefore represent two fundamentally different pathways by which animal chromosomes evolve, one refining existing architectures and the other rewiring them more drastically, together shaping the balance between stability and innovation in metazoan genome evolution.

The outcomes of this research open several promising directions for future exploration. Expanding chromosome-scale assemblies to additional cnidarian lineages, such as *Clytia hemisphaerica*, *Podocoryna carnea*, and *Cladonema pacificum*, will enable the reconstruction of ancestral TE repertoires and test whether the patterns of TE-driven genome expansion observed in this study are broadly shared across hydrozoans. Integrating long-read sequencing with single-cell multi-omics technologies will allow direct assessment of TE regulation and activation in specific stem-cell contexts, providing insight into how chromatin state influences TE behavior and how TE activity, in turn, shapes chromatin architecture. Comparative approaches between *Hydra* and *Hydractinia* will further clarify how differences in stem-cell potency and life-history strategies have co-evolved with genome evolution mediated by TE activity. Ultimately, this work underscores the importance of viewing TEs not as genomic anomalies but as integral evolutionary forces that continue to shape both genome function and organismal evolution.

References

- Kon-Nanjo, K., T. Kon, T. C.-T. K. Yu, D. Rodriguez-Terrones, F. Falcon *et al.*, 2025 The dynamic genomes of *Hydra* and the anciently active repeat complement of animal chromosomes. *Genome Biol.* 26: 186.
- Kon, T., K. Kon-Nanjo, and O. Simakov, 2025 Subtelomeric repeat expansion in *Hydractinia symbiolongicarpus* chromosomes. *Mob. DNA* 16: 14.
- Kon-Nanjo, K., T. Kon, H. R. Horkan, Febrimarsa, R. E. Steele *et al.*, 2023 Chromosome-level genome assembly of *Hydractinia symbiolongicarpus*. *G3 (Bethesda)*. 13.