Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus *Hydra*

Wai Yee Wong[a], Oleg Simakov[a,1], Diane M. Bridge[b], Paulyn Cartwright[c], Anthony J. Bellantuono[d], Anne Kuhn[e], Thomas W. Holstein[e], Charles N. David[f], Robert E. Steele[g], and Daniel E. Martínez[h,1]

[a]Department of Molecular Evolution and Development, University of Vienna, 1010 Vienna, Austria; [b]Department of Biology, Elizabethtown College, Elizabethtown, PA 17022; [c]Department of Ecology & Evolutionary Biology, University of Kansas, Lawrence, KS 66045; [d]Department of Biological Sciences, Florida International University, Miami, FL 33199; [e]Centre for Organismal Biology, Heidelberg University, 69120 Heidelberg, Germany; [f]Faculty of Biology, Ludwig Maximilian University of Munich, 80539 Munich, Germany; [g]Department of Biological Chemistry, University of California, Irvine, CA 92617; and [h]Department of Biology, Pomona College, Claremont, CA 91711

Edited by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved October 8, 2019 (received for review July 9, 2019)

Transposable elements are one of the major contributors to genome-size differences in metazoans. Despite this, relatively little is known about the evolutionary patterns of element expansions and the element families involved. Here we report a broad genomic sampling within the genus *Hydra*, a freshwater cnidarian at the focal point of diverse research in regeneration, symbiosis, biogeography, and aging. We find that the genome of *Hydra* is the result of an expansion event involving long interspersed nuclear elements and in particular a single family of the chicken repeat 1 (CR1) class. This expansion is unique to a subgroup of the genus *Hydra*, the brown hydras, and is absent in the green hydra, which has a repeat landscape similar to that of other cnidarians. These features of the genome make *Hydra* attractive for studies of transposon-driven genome expansions and speciation.

Transposable elements (TEs) were originally discovered by Barbara McClintock in maize (1) and later found to comprise a significant fraction of plant and animal genomes (2). Well-known for their contribution to total genome size (most recently in refs. 3 and 4), transposons are also sources of regulatory element evolution, modulators of gene expression (5), and a potential basis of large-scale genomic rearrangements (6).

Hydra provides an intriguing system to study the evolutionary history of TEs. The genus is subdivided into 2 major groups: the brown hydras (comprised of the Vulgaris, Oligactis, and Braueri clades) and the algal symbiont-containing green hydra (comprised of the Viridissima clade) (7). Cnidarian genomes are typically smaller than 500 Mb in size (8), as in the Viridissima clade, which has a genome size of about 300 Mb. In contrast, the genomes of brown hydras are ∼1 Gb in size (9). High abundance of TEs in the *Hydra vulgaris* strain 105 genome (9) has led to the hypothesis that large genome size is due to their expansion in this taxon. However, genomic data from other *Hydra* lineages were required to determine the timing of TE expansion and to rule out other scenarios, such as genome duplication in brown hydras.

To address this question, we sequenced genomes and transcriptomes from 4 brown hydras and 1 *Hydra viridissima* strain (*Materials and Methods*). Using the *H. vulgaris* strain 105 gene set as a reference (10), we constructed single-ortholog gene families using mutual best BLAST (basic local alignment search tool) hits to a select set of species (Fig. 1*A*). We used RAxML (11) to construct a phylogeny (Fig. 1*A*). The branching order for the 4 *Hydra* clades was identical to that found previously (7). We used r8s (12) to estimate divergence times, setting the cnidarian–bilaterian divergence as a calibration point to 550 Mya (13). We obtained 87 Mya for the beginning of the *Hydra* radiation and 59 Mya for the timing of the brown/green hydra split. These estimates for *Hydra* radiation times are based on transcriptome data and fall between previously reported estimates (7, 14) (Fig. 1*B*).

Using transcriptome data, we searched for evidence of a genome duplication event in the brown hydras. We found that 75% (8,629 out of 11,543) of gene families had the same number of genes in both *H. viridissima* and *H. vulgaris*. Additionally, 84.7% and 81.1% of the gene families contained a single gene from *H. vulgaris* and *H. viridissima*, respectively. Thus, there was no evidence for genome duplication as the explanation for the large genome size in the brown hydras. However, a good reference assembly for the green hydra genome will be required to completely rule out a rediploidization scenario.

To test the contribution of TEs to the *Hydra* genome expansion, we used DNAPipeTE (15) to identify and assemble highly abundant DNA reads from a random sample of 1 million reads from each species. We found that all of the major TE classes are represented at similar levels in the *Hydra* genomes with the exception of long interspersed nuclear elements (LINEs). LINEs were strikingly enriched (>6-fold) in the brown hydra genomes (Fig. 1*B*). We found that 2 of the 3 major LINE classes are overrepresented, L2 and CR1 (16), comprising at least 8 to 12% of all sampled reads in brown hydras, compared to less than 0.5% in the green hydra.

To determine the evolutionary history of LINEs in the brown hydras, we constructed a similarity graph based on BLASTN scores among all detected LINE consensus sequences. While we could identify contributions from all CR1/L2 families, we found that the expansion was largely limited to a specific region of the graph (Fig. 2). This indicates that the largest expansion happened in only one or a few highly related CR1 families (as defined by DNAPipeTE), together responsible for at least 28% of the expansion among the brown hydras.

We next investigated whether the CR1 expansion happened independently in each brown hydra lineage or at the base of the brown hydra clade. Based on similarity graphs of CR1 families (Fig. 2), we found that more than half of all DNAPipeTE CR1 families (e.g., 85 out of 116 CR1 families in *Hydra circumcincta* and 60 out of 105 CR1 families in *Hydra oligactis*) could be traced

Author contributions: W.Y.W. designed research; W.Y.W., O.S., D.M.B., P.C., A.J.B., A.K., T.W.H., C.N.D., R.E.S., and D.E.M. performed research; W.Y.W., O.S., and D.E.M. analyzed data; and W.Y.W., O.S., D.M.B., P.C., A.J.B., A.K., T.W.H., C.N.D., R.E.S., and D.E.M. wrote the paper.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: Sequences have been deposited in the NCBI Sequence Read Archive entry for the project (PRJNA114713). The full analysis pipeline is available from https://github.com/nicew/hydracompgen.

[1]To whom correspondence may be addressed. Email: oleg.simakov@univie.ac.at or dem04747@pomona.edu.

First published October 28, 2019.

Downloaded by guest on June 12, 2020

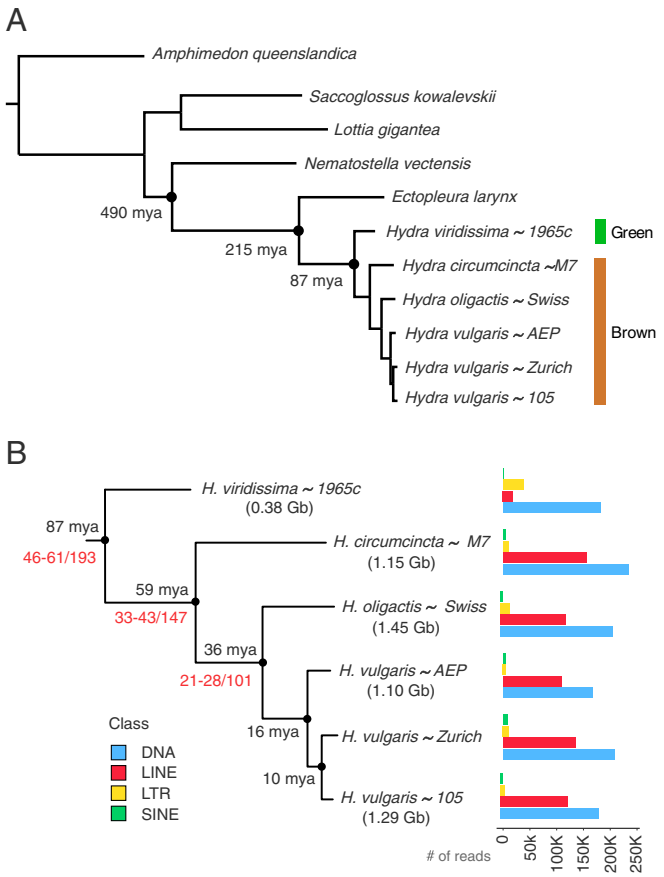

Fig. 1. Phylogeny of the genus *Hydra*. (*A*) RAxML phylogeny of *Hydra* and other metazoans. Dating estimates are provided for each node. (*B*) Branching pattern and divergence within the genus *Hydra*. Age estimates in black are from this study, those in blue are from ref. 7, and those in red are from refs. 14 and 24. Bar charts on the right indicate proportions of the major repeat element classes in each species. Genome sizes are from refs. 25 and 26.

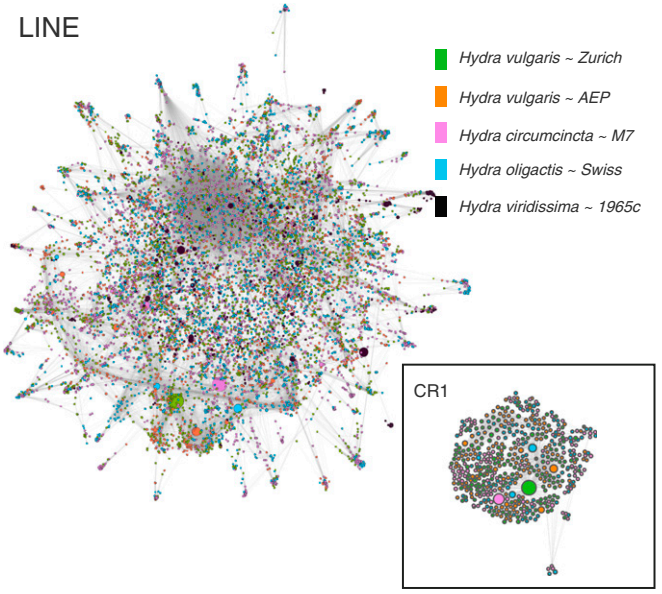

Fig. 2. Expansion of a specific CR1 retrotransposon family in the brown hydra lineage. Sequence similarity graph of LINE families for 5 *Hydra* species. The sizes of the circles correspond to the number of elements within each family, with species colored according to the key. Expansion is restricted to only one part of the graph (larger circles), stemming mostly from a single retrotransposon family (CR1, expanded panel).

back to the last common brown hydra ancestor. Interestingly, the majority of CR1 sequences in the genomes apparently lack the ability to propagate autonomously, as they are relatively short (478 bp on average, estimated by RepeatCraft, ref. 17).

Taken together, our findings show that a single CR1 family dominated the CR1/L2 LINE expansion after the separation of the green and brown hydra lineages. Given the wide distribution of these elements across the genome, an alternative scenario of repetitive element excision events that happened only in the green hydra lineage seems unlikely. Moreover, the repeat content of the green hydra genome is similar to that of other cnidarians, for example *Nematostella vectensis* (18), suggesting that this was the ancestral state for *Hydra* and for the phylum Cnidaria in general.

The observed genome expansion pattern in the genus *Hydra* is strikingly different from the other reported recent expansions, such as in larvaceans (3) and rotifers (4), in which expansion is due to a combination of various repeat element classes. These observations highlight the diversity of genome expansion events and demonstrate the importance of carrying out genome comparisons across a taxonomic group. Our results also indicate that *Hydra* will be an attractive model system for a targeted study of repeat-driven genome expansion and the role of repeat expansions in speciation.

Materials and Methods

Hydra cultures were maintained using standard methods (19). DNA and RNA extractions were done using Qiagen kits. Library preparation and Illumina sequencing were done using standard methods. Sequences have been deposited in the NCBI Sequence Read Archive entry for the project (PRJNA114713). Transcriptomes were assembled with Trinity (20), filtered with CD-hit (21), and the peptides predicted with Transdecoder (22). Orthologous groups were constructed using OrthoFinder (23). The full analysis pipeline is available from https://github.com/niccw/hydracompgen.

ACKNOWLEDGMENTS. This work was made possible, in part, through access to the Genomics High Throughput Facility Shared Resource of the Cancer Center Support Grant P30CA-062203 at the University of California, Irvine; and NIH shared instrumentation grants 1S10RR025496-01, 1S10OD010794-01, and 1S10OD021718-01. W.Y.W. and O.S. were supported by Austrian Science Fund Grant P30686-B29. D.E.M. and D.M.B. were supported by National Institute on Aging Grant 1R01AG037965-01 and a Templeton Foundation Immortality Project grant. D.E.M. was additionally supported by the Pomona College Sara and Egbert Schenck Memorial Fund. T.W.H. was supported by German Science Foundation Grants SFB-873-A1-3 and SFB-1324-A5-1. R.E.S. was supported by funds from the University of California, Irvine, School of Medicine. P.C. was supported by the National Science Foundation CAREER Award DEB 0953571. We thank Dr. A. Nawrocki for help in obtaining the *Ectoplura larynx* transcriptome.

1. B. McClintock, The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.* **36**, 344–355 (1950).
2. A. Canapa, M. Barucca, M. A. Biscotti, M. Forconi, E. Olmo, Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* **147**, 217–239 (2015).
3. M. Naville *et al.*, Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr. Biol.* **29**, 1161–1168.e6 (2019).
4. J. Blommaert, S. Riss, B. Hecox-Lea, D. B. Mark Welch, C. P. Stelzer, 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).
5. C. Feschotte, Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405 (2008).
6. J. K. Lim, M. J. Simmons, Gross chromosome rearrangements mediated by transposable elements in Drosophila melanogaster. *BioEssays* **16**, 269–275 (1994).
7. D. E. Martínez *et al.*, Phylogeny and biogeography of *Hydra* (Cnidaria: Hydridae) using mitochondrial and nuclear DNA sequences. *Mol. Phylogenet. Evol.* **57**, 403–410 (2010).
8. K. Adachi, H. Miyake, T. Kuramochi, K. Mizusawa, S. Okumura, Genome size distribution in phylum Cnidaria. *Fish. Sci.* **83**, 107–112 (2017).

Downloaded by guest on June 12, 2020

9. J. A. Chapman *et al*., The dynamic genome of *Hydra*. *Nature* **464**, 592–596 (2010).
10. National Human Genome Research Institute, *Hydra* 2.0 genome Project Portal. https://research.nhgri.nih.gov/hydra/. Accessed 18 October 2019.
11. A. Stamatakis, T. Ludwig, H. Meier, RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**, 456–463 (2005).
12. M. J. Sanderson, r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
13. L. A. Parry *et al*., Ichnological evidence for meiofaunal bilaterians from the terminal Ediacaran and earliest Cambrian of Brazil. *Nat. Ecol. Evol.* **1**, 1455–1464 (2017).
14. M. Schwentner, T. C. G. Bosch, Revisiting the age, evolutionary history and species level diversity of the genus *Hydra* (Cnidaria: Hydrozoa). *Mol. Phylogenet. Evol.* **91**, 41–55 (2015).
15. C. Goubert *et al*., De novo assembly and annotation of the Asian tiger mosquito (Aedes albopictus) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (Aedes aegypti). *Genome Biol. Evol.* **7**, 1192–1205 (2015).
16. C. J. Metcalfe, D. Casane, Modular organization and reticulate evolution of the ORF1 of Jockey superfamily transposable elements. *Mob. DNA* **5**, 19 (2014).
17. W. Y. Wong, O. Simakov, RepeatCraft: A meta-pipeline for repetitive element defragmentation and annotation. *Bioinformatics* **35**, 1051–1052 (2019).
18. N. H. Putnam *et al*., Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
19. H. M. Lenhoff, R. D. Brown, Mass culture of hydra: An improved method and its application to other aquatic invertebrates. *Lab. Anim.* **4**, 139–154 (1970).
20. M. G. Grabherr *et al*., Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
21. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
22. B. J. Haas *et al*., De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
23. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
24. A. M. Nawrocki, A. G. Collins, Y. M. Hirano, P. Schuchert, P. Cartwright, Phylogenetic placement of *Hydra* and relationships within Aplanulata (Cnidaria: Hydrozoa). *Mol. Phylogenet. Evol.* **67**, 60–71 (2013).
25. H. Zacharias, B. Anokhin, K. Khalturin, T. C. G. Bosch, Genome sizes and chromosomes in the basal metazoan Hydra. *Zoology (Jena)* **107**, 219–227 (2004).
26. G. Hemmrich, B. Anokhin, H. Zacharias, T. C. G. Bosch, Molecular phylogenetics in Hydra, a classical model in evolutionary developmental biology. *Mol. Phylogenet. Evol.* **44**, 281–290 (2007).

Downloaded by guest on June 12, 2020