# Accumulation of a biparentally-inherited Neptune transposable element in natural Killifish hybrids (*Fundulus diaphanus* × *F. heteroclitus*)

Alexandre-James Roussel,[1] Alexander Suh,[2,3,4,5] Francisco J. Ruiz-Ruano[2,3,4*], Anne-Marie Dion-Côté[1*],

[1]Département de biologie, Université de Moncton, Moncton, NB, Canada

[2]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, United Kingdom

[3]Centre for Molecular Biodiversity Research, Leibniz Institute for the Analysis of Biodiversity Change, Bonn, Germany

[4]Bonn Institute for Organismal Biology – Animal Biodiversity, University of Bonn, Bonn, Germany

[5]Department of Organismal Biology – Systematic Biology (EBC), Science for Life Laboratory, Uppsala University, Uppsala, Sweden

*Corresponding authors: Anne-Marie Dion-Côté, Département de biologie, Université de Moncton, Moncton, NB, Canada, +1-506-858-4815, anne-marie.dion-cote@umoncton.ca; Francisco J. Ruiz-Ruano, Centre for Molecular Biodiversity Research, Leibniz Institute for the Analysis of Biodiversity Change, Adenauerallee 127, 53113 Bonn, Germany, F.RuizRuano@leibniz-lib.de

## Abstract

Transposable elements (TEs) are abundant selfish genetic elements that can mobilize in their host genome, causing DNA damage, mutations and chromosome rearrangements. TE silencing is thus critical, and is initiated by maternally loaded piRNAs, leading to their repression. Consistently, paternally inherited TEs are derepressed in the progeny of Drosophila crosses involving a naive female. TEs have also been found to be derepressed in interspecific crosses, which is proposed to result from suboptimal interactions of piRNA pathway proteins. *Fundulus heteroclitus* and *F. diaphanus* hybridize in nature and produce viable and fertile offspring that sometimes reproduce asexually. We characterized the repetitive DNA content of these species and their asexually reproducing hybrids. TE load was slightly higher than expected in hybrids and associated with younger repeats. A bi-parentally inherited active Neptune element showed a remarkable ~4-fold accumulation in hybrids. These results are consistent with suboptimal piRNA pathway function, leading to active TE accumulation.

**Keywords:** Repetitive DNA, transposable element, hybridization, asexual reproduction, fish

## Introduction

Transposable elements (TEs) are selfish repetitive sequences that can mobilize within their host genome. Class I elements (retrotransposons) require an RNA intermediate which is retro-transcribed into complementary DNA before re-inserting in the genome, increasing copy number. Retrotransposons are further classified into subclasses, depending on the presence or absence of long terminal repeats (LTRs and non-LTRs) (Wicker et al. 2007). *Penelope*-like elements (PLEs) form an additional subclass due to their unique organisation and mode of mobilization, and the presence of a unique GIY-YIG endonuclease domain (Evgen'ev and Arkhipova 2005; Frangieh et al. 2024). Class II elements (DNA transposons) do not transpose by reverse transcription but rather via DNA excision and insertion (Wicker et al. 2007).

TE mobilization can be highly deleterious, disrupting genes or regulatory elements, and inducing DNA damage or chromosome rearrangements (Bourque et al. 2018). Eukaryotes are thus equipped with silencing mechanisms to repress TEs (Cosby et al. 2019). In animals, a central component of this repression system is composed of the Piwi-interacting RNAs (piRNA) and the piRNA pathway proteins, which lead to TE heterochromatinization and repression (Le Thomas et al. 2013). This silencing is initiated by maternally transmitted piRNAs, safeguarding the progeny's genome against TEs (Malone et al. 2009). Thus, paternally-transmitted active TEs not carried by the maternal line can be derepressed and mobilize in hybrids, such as in so-called dysgenic crosses of Drosophila (Bucheton et al. 1976; Kidwell et al. 1977; Evgen'ev et al. 1997).

The sequences of TEs and piRNA pathway proteins evolve rapidly, which is at least partly driven by the rapid evolution of selfish TEs, akin to an evolutionary arms race (Blumenstiel et al. 2016; Cosby et al. 2019; Parhad and Theurkauf 2019). In contrast to dysgenics crosses, both maternally and paternally transmitted TEs are derepressed in

Drosophila interspecific hybrids (Labrador et al. 1999; Kelleher et al. 2012; Romero-Soriano et al. 2017) and some mammal, plant and fish hybrids (O'Neill et al. 1998; Ungerer et al. 2006; Dion-Côté et al. 2014). This led Kelleher *et al.* (2012) to propose that piRNA pathway proteins accumulate genetic incompatibilities over time, leading to bi-parentally inherited TE derepression in inter-specific hybrids.

*Fundulus heteroclitus* and *F. diaphanus* are killifish species that diverged between 15 and 25 million years ago (Ghedotti and Davis 2017). While *F. diaphanus* is primarily found in freshwater, *F. heteroclitus* preferentially occurs in saltwater (Scott and Crossman 1973). Both species are euryhaline and co-occur in brackish water, where sexually and asexually reproducing hybrids have been reported (Fritz and Garside 1974; Dawley 1992; Hernández Chávez and Turgeon 2007; Mérette et al. 2009). An opening formed ~70 years ago (~25-40 generations) in Porters Lake, NS, allowing sea water to enter (Fritz and Garside 1974). Therefore, secondary contact between *F. heteroclitus* and *F. diaphanus* in Porters Lake is very recent, from ~70 years to a conservative maximum of 13 000 years, as the area was covered by Wisconsinian glaciers until then (Shaw et al. 2006; Dalziel et al. 2020). There, previous work has revealed the presence of asexually reproducing "frozen F1" hybrids, most likely reproducing by gynogenesis (Hernández Chávez and Turgeon 2007; Dalziel et al. 2020). Several clonal lineages co-exist, but one lineage is predominant (Hernández Chávez and Turgeon 2007; Dalziel et al. 2020).

Herein, we first characterize and quantify repetitive DNA content in *F. diaphanus* and *F. heteroclitus*. We then test whether some maternally and/or paternally transmitted TEs accumulated in their asexually reproducing "frozen F1" hybrids. The relatively young age of the *Fundulus* clones allows to capture TE accumulation in these hybrids, if any, before their decay, as predicted under long-term asexual reproduction (Hickey 1982; Arkhipova and Meselson 2005; Dolgin and Charlesworth 2006).

## Material and Methods

*Fish collection and sampling*

Experimental procedures were approved by the Université de Moncton's animal care and use committee (19-02). Fishes were collected from Porters Lake, Nova Scotia (44°43′40″N, 63°18′00″W) during Summer 2020 using beach seine nets under a Department of Oceans and Fisheries licence (355651). Because asexually reproducing hybrids are all females, only females were kept (Table S1). Fish were housed in brackish water (12 ppt) at Université de Moncton until they were sacrificed. Fin clips were preserved in 95% ethanol.

*DNA extraction, genotyping and sequencing*

DNA from three *F. heteroclitus*, four *F. diaphanus* and seven hybrid individuals was extracted from fin clips using the E.Z.N.A tissue DNA extraction kit (Omega Bio-tek). To determine maternal ancestry, a 441bp fragment from the mtDNA D-Loop was amplified by PCR and digested with HphI (New England BioLabs), as previously described (Dalziel et al. 2020). In addition, five microsatellites loci were genotyped for each individual [FhATG-B103, FhCA-1 and FhCA-21: (Adams et al. 2005), Fhe113 and Fhe57: (Hernández Chávez and Turgeon 2007), Table S2]. Allelic sizes were determined by capillary electrophoresis on a 3730xl DNA Analyzer (Applied Biosystems) at The Center for Applied Genomics, Toronto. Results were analyzed using the Microsatellite Analysis application (ThermoFisher Connect Cloud).

PCR-free DNA libraries (Lucigen) were prepared and sequenced on an Illumina HiSeq Instrument with 150 bp paired-end reads at the Genome Quebec Innovation Center (Table S3). Quality control was performed using FastQC (V0.11.9). Reads were trimmed with Trimmomatic (v0.38) using the following parameters: ILLUMINACLIP 2:30:7:4, HEADCROP: 10 and MINLEN: 100.

*Repeat characterization*

We characterized the most abundant repeats in hybrid individuals in a single run of RepeatExplorer2 (Novák et al. 2013). We randomly selected and concatenated 71,429 trimmed read pairs from each of the seven hybrid individuals yielding a total of 500,000 read pairs. We classified clusters and superclusters into the most specific group possible based on a hierarchical decision tree (Fig. S1). “Unclassified repeats” are clusters that were not annotated as a satellite, mobile element or rDNA, whereas “Unknown” clusters were not classified as organelle or contamination. The final repeat database was composed of 14,011 contigs (Supplementary File 1).

We used the “id_rmasker.py” script to add the annotation to the sequence IDs with the RepeatMasker’s format (Smit et al. 2013-2015) and the “replace_patterns.py” script to transform the RepeatExplorer2 annotation to be cluster-specific (https://github.com/fjruizruano/ngs-protocols). We used the resulting repeat database to estimate the proportion of the genome occupied by each cluster. To do this, we selected 5 million trimmed read pairs per library using seqtk v1.3 (https://github.com/lh3/seqtk). We then quantified the abundance of the database repeats in this subset of reads using RepeatMasker v4.1.3 (Supplementary Files 2–3). We ran RepeatMasker as integrated in the repeat_masker_run_big.py script from the satMiner toolkit (https://github.com/fjruizruano/satminer, last access on 11/07/2022) (Ruiz-Ruano et al. 2016). This script runs RepeatMasker on large query files by splitting them up into subsets. It also runs the calcDivergenceFromAlign.pl script from RepeatMasker, which counts the total number of nucleotides occupied by each group of repeats along divergence intervals (Kimura substitution levels) of 1% needed to generate repeat landscapes. Genome proportion occupied by each cluster was obtained by normalizing the number of nucleotides occupied by each cluster to the total number of nucleotides for its subset of reads. We then generated average repeat landscapes for each group (*F. heteroclitus*, *F. diaphanus*, and hybrids) and subtractive repeat

landscapes as the values of a group minus the values of other group. A principal component analysis was run on the abundance for each repeat class using the function PCA() from the FactoMineR package (Lê et al. 2008) in R. We calculated the average genome proportion of the parental species and the hybrids for each cluster. We then detected important deviations from the observed genome proportion occupied by each repeat cluster in hybrids compared with the expected proportion as the average between the genome proportion of both parental species. This analysis revealed repeat.CL27 as being ~4X more abundant than expected.

*Haplotype calling*

We tested whether some species-specific sub-families of repeat.CL27 were preferentially accumulated in hybrids. For this, we used the ORF2 within the repeat.CL27 consensus sequence from RepeatExplorer as a reference (see below). We mapped the trimmed paired reads from the 14 individuals with bwa-mem2 v.2.2.1 using relaxed conditions ("-SP -k 14 -L 0,0 -U 0 -B 1"). We processed the resulting BAM files using samtools v1.14 (Danecek et al. 2021) as follows. First, we sorted the reads and kept mappings with MAPQ ≥ 40 and merged the mapped reads for the parental species. We then added the read group for each BAM file with Picard v3.1.1 (Broad Institute). We performed haplotype calling with freebayes v1.3.2 using a minimum length of 7 (--haplotype-length 7) (Garrison and Marth 2012). We extracted haplotypes from the resulting VCF file with BCFtools query v.1.10.2 (Danecek et al. 2021) applying the option (-f '%CHROM\t%POS\t%REF\t%ALT\t[%SAMPLE:%GT\t]\n'). Then, we extracted private haplotypes from each parental species and selected haplotypes longer than 15 nt. Finally, we extracted the mapped reads from the 14 libraries (parental species and hybrids) in FASTQ format with samtools fastq and counted the number of occurrences of each haplotype with grep (Supplementary File 4).

*Copy number validation by qPCR*

We identified open reading frames (ORF) in the consensus sequence of repeat.CL27 (CL27Contig27_13_sc_0.609135_l_5617) using ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/), which identified two long non-overlapping ORFs (ORF1 and ORF2, 420 and 2004 nt, respectively). We then estimated the number of copies in 18 individuals (six individuals for each group) by qPCR. For this, we used Primer-BLAST to design primers targeting the ORF2 from repeat.CL27 (F: 5'-AGGAACCTGAAAAGTGGAGCA-3', R: 5'-TGGTGCCGAATTGAGAGAGG3') and the citrate synthase gene from the *F. heteroclitus* reference genome (NC_046361, F: 5'-TGCACCCCATGTCTCAGTTC-3', R: 5'-TTGGACTTGTGCACTCCCTC-3).

Efficiency curves were generated using 0.0032 - 10 ng (5-fold serial dilution) of gDNA from a single *F. heteroclitus* female individual. Reactions (7 µl) contained 0.25 µM of each primer and 1X Luna Universal qPCR master mix (New England BioLabs). Amplification began with an initial denaturation at 95°C for 1 min, followed by denaturation at 95°C for 15 sec and combined annealing and amplification at 60°C for 30 sec, for a total of 40 cycles. Amplification efficiency was acceptable (97.3%, $R^2$ = 0.994 for citrate synthase, 100.65%, $R^2$ = 0.994 for repeat.CL27 ORF2) and melting curves revealed a single peak. Both reactions were run on the same plate for each sample.

*Cloning and sequencing of repeat.CL27 ORF2*

We manually designed primers to amplify the full-length reverse transcriptase encoded by ORF2 from repeat.CL27 (F: 5'-ATGAGGTATGACCTGCAAAAATACC-3', R: 5'-GATCCAAATCCTTTCAGCCCTTC-3'). The reactions (25 µl) contained 0.6 µM of each primer, 2 mM $MgCl_2$, 2 mM dNTPs, 1.25 U GoTaq and 45 ng of DNA. The amplification protocol included an initial denaturation step at 95°C for 2 min, followed by 32 cycles of denaturation at 95°C for 1 min, annealing at 60°C for 1 min and amplification at 72°C for 1 min, with a final

elongation at 72°C for 5 min. Amplicons were gel-purified using the Monarch DNA gel extraction kit (New England BioLabs). Purified amplicons were cloned using the TOPO TA cloning kit for sequencing (Invitrogen) following the manufacturer's recommendations and transformed in DH5 competent cells (Fisher Scientific). Sequencing of the plasmid was performed by Plasmidsaurus, and vector sequences were removed using A Plasmid Editor (Davis and Jorgensen 2022) (Supplementary file 5).

## Results and Discussion

Analysis of the parental species revealed that *F. diaphanus* had a higher repetitive DNA load than *F. heteroclitus* (30.39% vs 29.30%, Table S4), consistent with its larger genome (1.48 Gb for *F. diaphanus* vs 1.33 Gb for *F. heteroclitus*, (Dawley 1992)). For *F. heteroclitus*, the proportion of the genome occupied by TEs was slightly lower than previously reported using an assembly-based approach, which is more sensitive to low copy number repeats in the assembled genome (32.66%, (Reinar et al. 2023)). In fishes, TEs are highly diversified, with non-LTR retrotransposons and DNA transposons typically representing the most abundant classes, many of which remain active (Volff 2005; Shao et al. 2019). As expected, non-LTR retrotransposons were the most abundant TEs in both species. We identified very few DNA transposons (<1% of the genome for Maverick elements and no hATs), but this would be expected if many of them were non-autonomous, i.e., lacking any ORFs (Peona et al. 2024). We compared the repetitive DNA content of *F. diaphanus* and *F. heteroclitus* by generating repeat landscapes which revealed higher abundances among the lower Kimura substitution level bins for both species, consistent with a recent accumulation of repeats (Fig. 1A & B). A subtractive repeat landscape, calculated as *F. diaphanus* minus *F. heteroclitus*, revealed that the main differences between both species were found among the less divergent repeats, i.e., those that expanded more recently (Fig. 1C).

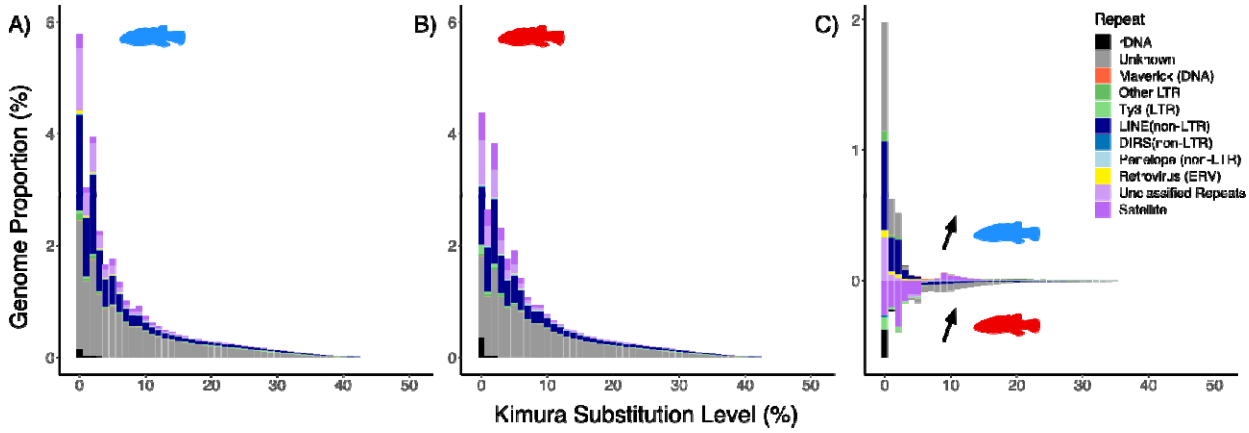


**Figure 1.** Repeat content in *Fundulus* spp. We calculated the genome proportion occupied by each repeat group as a function of the Kimura substitution level (CpG corrected) of a given repeat cluster to its consensus sequence. A) Repeat landscape of *F. diaphanus*. B) Repeat landscape of *F. heteroclitus*. C) Subtractive repeat landscape generated by subtracting *F. heteroclitus* (B) from *F. diaphanus* (A) abundance values. Annotation based on Fig. S1. Silhouette from PhyloPic.

To visualize how repeat content in hybrids differs from the parental species, we performed a principal component analysis (PCA) on the genome proportion occupied by each repeat group. Principal components (PC) 1 (66.1% of the variance) clearly separated parental species, which was expected given their contrasted repeat landscapes (Fig. 2A). Along PC1, hybrids were also intermediate between parental species, although slightly closer to *F. diaphanus*. This was expected as *F. diaphanus* has a larger genome. PC2 (25.4% of the variance) clearly separates hybrids from parental species, with all repeat groups being positively correlated to this PC, with the exception of rDNA and retroviruses. This suggests that most repeat groups accumulated in hybrid genomes. We thus generated a subtractive repeat landscape to compare the observed hybrid repeat content with the “expected” repeat content (i.e., the sum of half the average abundance of each parental species for each repeat group, Fig. 2B). This revealed that hybrids had a higher repetitive DNA content than expected for every repeat group, except rDNA (Fig. 2C). Yet, the genome proportion occupied by repeats was only 0.65% higher in hybrids than in either parental species (Table S4). These results suggest that most classes of repetitive sequences have accumulated in *Fundulus* hybrids, consistent with

suboptimal piRNA pathway protein interaction for TEs, although very moderately. While we cannot entirely exclude technical biases, such as underestimating hybrid genome size or overestimating repeat abundance through parental allele separation, this pattern would not be restricted to the youngest repeats (i.e. lowest Kimura substitution bins) as we observe.

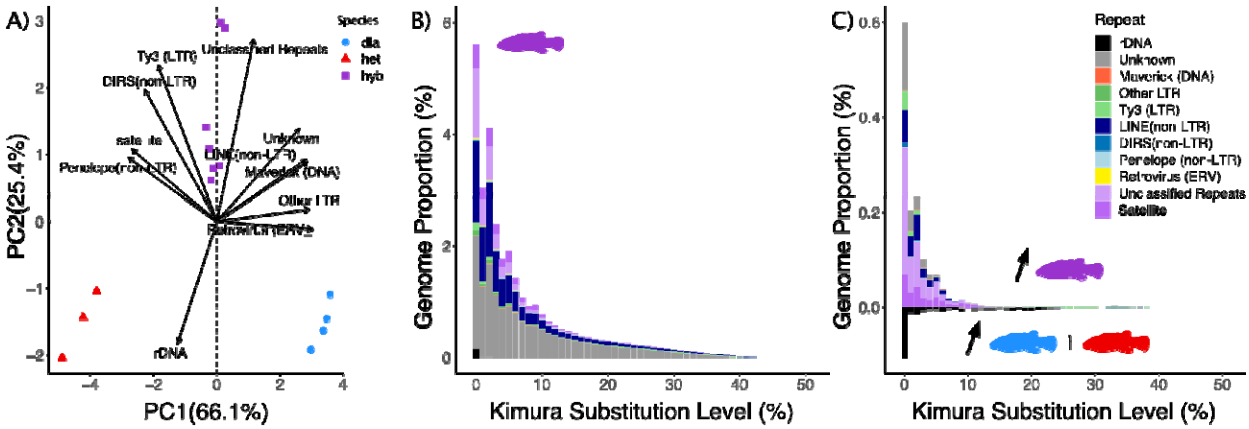


**Figure 2.** Repeat content in *Fundulus* hybrids. A) PCA of the repeat class content on *Fundulus* fish. B) Repeat landscape of *Fundulus* hybrids. C) Subtractive repeat landscape generated by subtracting expected abundance values (half the abundance from *F. diaphanus* and *F. heteroclitus*) from the abundance observed in hybrids. Annotation based on Fig S1.

Therefore, we examined whether some repeats accumulated disproportionately in hybrids. For this, we extracted the only three repeat clusters that varied by more than 1 Mb between the observed and expected hybrid repeat content. The first one, annotated as a Ty3 element (LTR retrotransposon), was ~1.15 times more abundant in hybrids than expected, indicating a moderate increase. The second one, satellite.CL1, fell within the *F. heteroclitus* range. A third candidate, repeat.CL27 (“unclassified”, Fig. S1), stood out, being ~4 times more abundant than expected in hybrids (Fig. 3A, B). This result was further validated by qPCR with different individuals, to exclude methodological biases (Fig. S2). Based on CENSOR (Kohany et al. 2006), this repeat presents similarity (positions 5128 to 2560, minus strand) with a Neptune retroelement previously identified in the Threespine Stickleback (*Gasterosteus aculeatus*) (Neptune-2_GA, positions 5128-2460, 66.03% identity). Neptune elements are poorly

characterized *Penelope*-like elements defined by the presence of a reverse transcriptase (RT) and a GIY-YIG endonuclease domain (Evgen'ev and Arkhipova 2005) that have now been detected across algae, plants, fungi, protists and animals (Craig et al. 2021). We identified two long ORFs in repeat.CL27, from position 73-492 (419 nt, ORF1) and from position 666-2669 (2003 nt, ORF2). We were not able to detect sequence similarity for ORF1 in any of the queried databases (Repbase, GenBank, gtRNAdb, CDD), but ORF2 contains the intact RT domain (AA 590-558) and GIY-YIG (AA 299-392) domains based on the Conserved Domains Database (CDD). A phylogeny with other PLEs revealed the clustering of ORF2 with other Neptune elements (Fig. 3C). We cloned and sequenced full ORF2 from various individuals, which revealed the existence of two subfamilies of this Neptune element that are specific to each parental species (Fig. 3D). Both of these subfamilies were found in hybrids, which prompted us to examine the accumulation of short, species-specific haplotypes in repeat.CL27 from the sequencing reads (Supplementary File 4). We detected 8 regions with 27 haplotypes of at least 15 nt in length, with 12 being specific to *F. diaphanus* and 7 to *F. heteroclitus* (i.e., short haplotypes that are diagnostic of a subfamily that is restricted to each parental species). Among them, 5 haplotypes from *F. diaphanus* showed a higher abundance in hybrids than expected (expected/observed median = 6.79), while 5 from *F. heteroclitus* were more abundant than expected (median = 7.12). Thus, some species-specific haplotypes are present in excess in asexually reproducing hybrids, whereas others are under-represented. These results suggest that species-specific Neptune elements from each parental species are accumulating in *Fundulus* hybrids, but that this effect is not widespread among all insertions of this element. We propose that accumulation may be restricted to currently active insertions from both parental species.

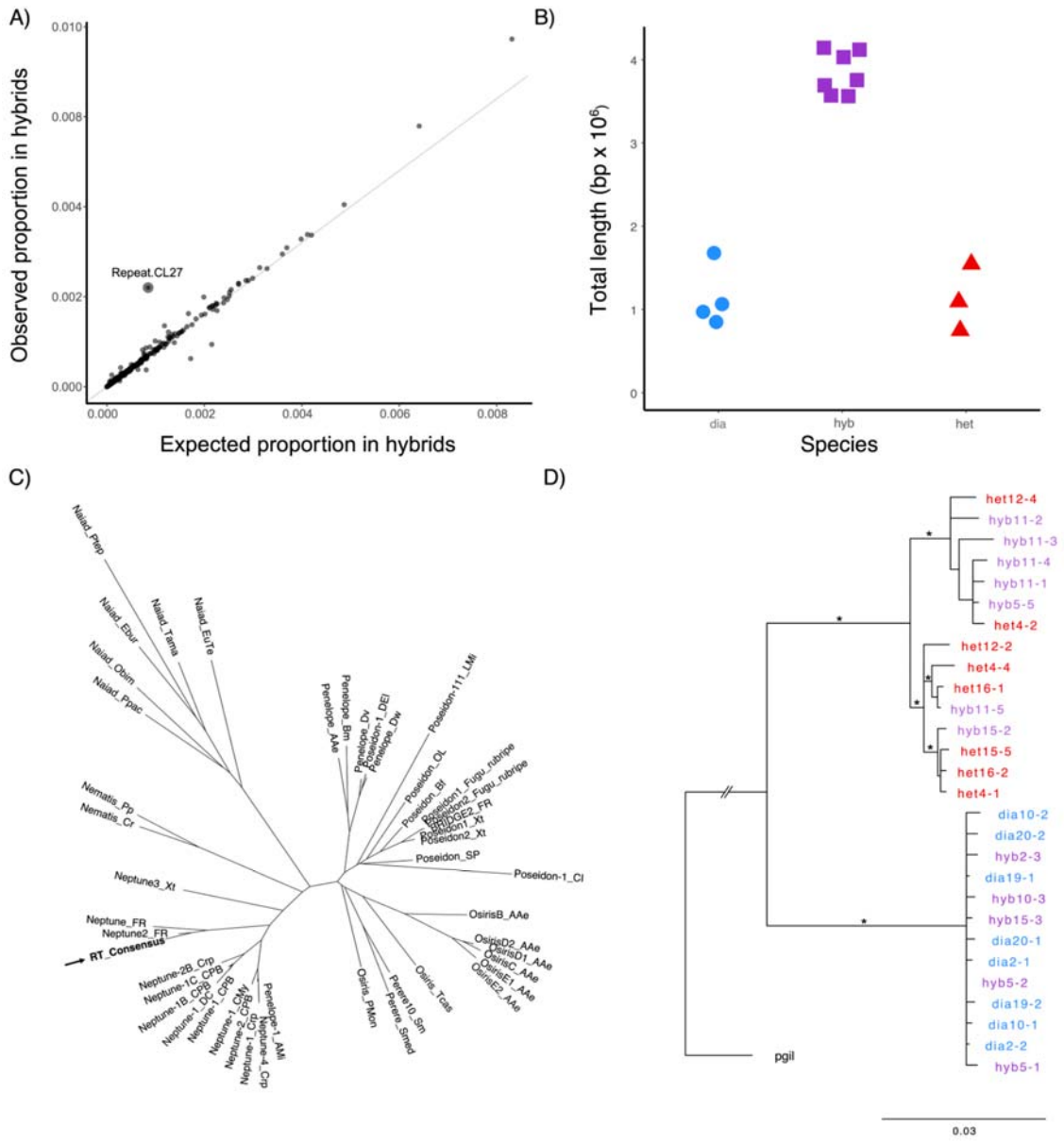


**Figure 3.** A Neptune transposable element accumulates in *Fundulus* hybrids. A) Observed proportion of the genome occupied by each repeat cluster in hybrids as a function of the expected copy number. Repeat.CL27 is highlighted. The line depicts a 1:1 ratio. B) Total genomic length (bp) occupied by repeat.CL27 based on sequencing. C) Phylogeny of PLEs RT and repeat.CL27 ORF2 (RT consensus). PLEs sequences were graciously provided by Tomás Carrasco-Valenzuela (Carrasco-Valenzuela et al. 2023). D) Phylogeny of cloned full-length RT from 14 individuals. Outgroup sequence from *Poecilia gillii* (NCBI accession number CAEGAJ010041810.1: 443 - 2446) (van Kruistum et al. 2021) was obtained by doing a BLASTN search of the consensus sequence against the WGS database restricted to the *Poecilia* genus (taxid:8080). Sequences were aligned using MAFFT-LINSI (v7.505) (REF, Supplementary file 6). We manually inspected the resulting alignment for the presence of PCR chimeras and removed them. We built a phylogeny using IQ-TREE v2.4.0 with default parameters and 1000

bootstraps (Nguyen et al. 2015). Trees were edited using Figtree v1.4.4 (https://github.com/rambaut/figtree). Branches with >0.85 bootstrap support are indicated with an asterisk.

## Conclusion

While the moderate accumulation of many repetitive sequences we observed in hybrids is coherent with suboptimal piRNA pathway protein interactions (Kelleher et al. 2012), technical biases cannot be entirely excluded. We also identified two subfamilies of a Neptune element that are specific to either *F. diaphanus* and *F. heteroclitus*, both of which accumulated dramatically in naturally occurring hybrids. This high accumulation of a single, biparentally inherited TE is in stark contrast to what is observed in Drosophila dysgenic crosses, whereby paternally-inherited TEs are derepressed, or other inter-specific crosses, in which multiple TEs accumulate (e.g., (Kidwell et al. 1977; Labrador et al. 1999; Kelleher et al. 2012)). The strong accumulation of this Neptune element suggests that this element is particularly aggressive. Because naturally occurring sexually and asexually reproducing hybrid lineages have repeatedly evolved, and can be compared with new lineages produced in the lab, we further propose that the *Fundulus* complex is an excellent system in which to disentangle hybridization-driven and asexuality-mediated effects on TE dynamics in the future.

## Data availability

Raw reads are available on SRA under PRJNA862733.

## Funding sources

This work was supported by a NSERC discovery grant to AMDC (RGPIN-2019-05744) as well as a New Brunswick Innovation Foundation Research Assistantship (RAI-0000000144) to AMDC. AJR was supported by a MITACS Globalink scholarship (IT32322). FJRR was supported by a Marie Curie Individual Fellowship (875732).

## Acknowledgements

We would like to thank Daniel Barbash, Mathieu Hénault, Karel Janko and France Dufresne for providing critical feedback. Part of the computations were enabled by resources in project NAISS 2024/22-720 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX, funded by the Swedish Research Council through grant agreement no. 2022-06725.

## References


Adams SM, Oleksiak MF, Duvernell DD. 2005. Microsatellite primers for the Atlantic coastal killifish, *Fundulus heteroclitus*, with applicability to related *Fundulus* species. *Mol. Ecol. Notes* 5:275–277.

Arkhipova I, Meselson M. 2005. Deleterious transposable elements and the extinction of asexuals. *Bioessays* 27:76–85.

Blumenstiel JP, Erwin AA, Hemmer LW. 2016. What drives positive selection in the Drosophila piRNA machinery? The genomic autoimmunity hypothesis. *Yale J. Biol. Med.* 89:499–512.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018. Ten things you should know about transposable elements. *Genome Biol.* 19:199.

Broad Institute. Picard. Available from: https://broadinstitute.github.io/picard/

Bucheton A, Lavige JM, Picard G, L'Heritier P. 1976. Non-mendelian female sterility in *Drosophila melanogaster*: quantitative variations in the efficiency of inducer and reactive strains. *Heredity (Edinb.)* 36:305–314.

Carrasco-Valenzuela T, Marins L, Ramos EKS, Suh A, Mazzoni CJ. 2023. Recent expansion of Penelope-like retrotransposons in the leatherback turtle Dermochelys coriacea. *Research Square* [Internet]. Available from: https://assets-eu.researchsquare.com/files/rs-2864407/v1/50317a00-a9e1-4f38-b561-8991674f018d.pdf?c=1700836693

Cosby RL, Chang NC, Feschotte C. 2019. Host–transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* [Internet]. Available from: http://genesdev.cshlp.org/content/33/17-18/1098.short

Craig RJ, Yushenova IA, Rodriguez F, Arkhipova IR. 2021. An ancient clade of Penelope-like retroelements with permuted domains is present in the Green lineage and protists, and dominates many invertebrate genomes. *Mol. Biol. Evol.* 38:5005–5020.

Dalziel AC, Tirbhowan S, Drapeau HF, Power C, Jonah LS, Gbotsyo YA, Dion-Côté A-M. 2020. Using asexual vertebrates to study genome evolution and animal physiology: Banded (*Fundulus diaphanus*) x Common Killifish (*F. heteroclitus*) hybrid lineages as a model system. *Evol. Appl.* 13:1214–1239.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* [Internet] 10. Available from: http://dx.doi.org/10.1093/gigascience/giab008

Davis MW, Jorgensen EM. 2022. ApE, A Plasmid Editor: A freely available DNA manipulation and visualization program. *Front. Bioinform.* 2:818619.

Dawley RM. 1992. Clonal hybrids of the common laboratory fish *Fundulus heteroclitus*. *Proc. Natl. Acad. Sci. U. S. A.* 89:2485–2488.

Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic

shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Mol. Biol. Evol.* 31:1188–1199.

Dolgin ES, Charlesworth B. 2006. The fate of transposable elements in asexual populations. *Genetics* 174:817–827.

Evgen'ev MB, Arkhipova IR. 2005. Penelope-like elements--a new class of retroelements: distribution, function and possible evolutionary significance. *Cytogenet. Genome Res.* 110:510–521.

Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, Corces VG. 1997. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis. *Proc. Natl. Acad. Sci. U. S. A.* 94:196–201.

Frangieh CJ, Wilkinson ME, Strebinger D, Strecker J, Walsh ML, Faure G, Yushenova IA, Macrae RK, Arkhipova IR, Zhang F. 2024. Internal initiation of reverse transcription in a Penelope-like retrotransposon. *Mob. DNA* 15:12.

Fritz ES, Garside ET. 1974. Identification and description of hybrids of *Fundulus heteroclitus* and *F. diaphanus* (Pisces: Cyprinodontidae) from Porters Lake, Nova Scotia, with evidence for absence of backcrossing. *Can. J. Zool.* 52:1433–1442.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]* [Internet]. Available from: http://arxiv.org/abs/1207.3907

Ghedotti MJ, Davis MP. 2017. The taxonomic placement of three fossil Fundulus species and the timing of divergence within the North American topminnows (Teleostei: Fundulidae). *Zootaxa* 4250:577–586.

Hernández Chávez C, Turgeon J. 2007. Asexual and sexual hybrids between *Fundulus diaphanus* and *F. heteroclitus* in the Canadian Atlantic region. *Mol. Ecol.* 16:1467–1480.

Hickey DA. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* 101:519–531.

Kelleher ES, Edelman NB, Barbash DA. 2012. Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biol.* 10:e1001428.

Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* 86:813–833.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7:474.

van Kruistum H, Nijland R, Reznick DN, Groenen MAM, Megens H-J, Pollux BJA. 2021. Parallel genomic changes drive repeated evolution of placentas in live-bearing fish. *Mol. Biol. Evol.* 38:2627–2638.

Labrador M, Farré M, Utzet F, Fontdevila A. 1999. Interspecific hybridization increases transposition rates of Osvaldo. *Mol. Biol. Evol.* 16:931–937.

Lê S, Josse J, Husson F. 2008. FactoMineR: AnRPackage for Multivariate Analysis. *J. Stat. Softw.* [Internet] 25. Available from: http://dx.doi.org/10.18637/jss.v025.i01

Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* 27:390–399.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* 137:522–535.

Mérette D, Bourret A, Turgeon J. 2009. Diversity and distribution of gynogenetic hybrids between *Fundulus diaphanus* and *Fundulus heteroclitus* in Porter's Lake (Nova Scotia) in relation to salinity. *J. Fish Biol.* 75:1115–1122.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268–274.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29:792–793.

O'Neill RJ, O'Neill MJ, Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393:68–72.

Parhad SS, Theurkauf WE. 2019. Rapid evolution and conserved function of the piRNA pathway. *Open Biol.* 9:180181.

Peona V, Martelossi J, Almojil D, Bocharkina J, Brännström I, Brown M, Cang A, Carrasco-Valenzuela T, DeVries J, Doellman M, et al. 2024. Teaching transposon classification as a means to crowd source the curation of repeat annotation - a tardigrade perspective. *Mob. DNA* 15:10.

Reinar WB, Tørresen OK, Nederbragt AJ, Matschiner M, Jentoft S, Jakobsen KS. 2023. Teleost genomic repeat landscapes in light of diversification rates and ecology. *Mob. DNA* 14:14.

Romero-Soriano V, Modolo L, Lopez-Maestre H, Mugat B, Pessia E, Chambeyron S, Vieira C, Garcia Guerreiro MP. 2017. Transposable element misregulation is linked to the divergence between parental piRNA pathways in Drosophila hybrids. *Genome Biol. Evol.* 9:1450–1470.

Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM. 2016. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* 6:28333.

Scott WB, Crossman EJ. 1973. Freshwater fishes of Canada: Fisheries Research Board of Canada Bulletin, Vol. 184. *Fisheries Research Board of Canada, Ottawa*.

Shao F, Han M, Peng Z. 2019. Evolution and diversity of transposable elements in fish genomes. *Sci. Rep.* 9:15399.

Shaw J, Piper DJW, Fader GBJ, King EL, Todd BJ, Bell T, Batterson MJ, Liverman DGE. 2006. A conceptual model of the deglaciation of Atlantic Canada. *Quat. Sci. Rev.* 25:2059–2081.

Smit AFA, Hubley R, Green P. 2013-2015. RepeatMasker Open-4.0. Available from: http://www.repeatmasker.org

Ungerer MC, Strakosh SC, Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol.* 16:R872–R873.

Volff J-N. 2005. Genome evolution and biodiversity in teleost fish. *Heredity (Edinb.)* 94:280–294.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8:973–982.