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. Two bi-parentally inherited active Neptune subfamilies showed a remarkable ~3-4-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

Hybridization between divergent populations or species may trigger genome instability in hybrids, a phenomenon long proposed to be associated with transposable element (TE) derepression (i.e., the genomic shock hypothesis, McClintock 1984). TEs are selfish repetitive sequences that can mobilize within their host genome and are classified in two major groups. 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 a peculiar type of non-LTR retrotransposons due to their unique organization 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). To counteract these threats, silencing mechanisms have evolved in eukaryotes to repress TEs (Cosby et al. 2019). In animals, a key component of this defense system involves Piwi-interacting RNAs (piRNA) and the associated piRNA pathway proteins, which together mediate TE repression (Le Thomas et al. 2013). In Drosophila, this silencing is initiated by maternally transmitted piRNAs, safeguarding the progeny's genome against TEs (Malone et al. 2009). Importantly, piRNA pathway proteins themselves evolve rapidly, a pattern that is widely interpreted as a response to the fast-paced sequence evolution of selfish TEs. This dynamic reflects an ongoing evolutionary arms race, where host silencing machinery must continually adapt to recognize and suppress

57 newly emerging or mutating TEs evading repression (Blumenstiel et al. 2016; Cosby et al. 2019;
58 Parhad and Theurkauf 2019).

59 Observations across various systems have led to two non-mutually exclusive models
60 that may account for TE mobilization in hybrids. The first model involves paternal-only TE
61 derepression, where active TEs transmitted by the paternal line but absent from the maternal
62 genome escape silencing and mobilize in hybrids. This phenomenon is thought to result from
63 the lack of maternally-transmitted piRNAs that match and silence paternally inherited TEs. This
64 phenomenon is well-documented in Drosophila dysgenic crosses but has not been thoroughly
65 descrived in other systems, to the best of our knowledge (Bucheton et al. 1976; Kidwell et al.
66 1977; Evgen'ev et al. 1997, Marie et al. 2017, Ryazansky et al. 2017). In contrast, the second
67 model involves bi-parental TE derepression, where TEs inherited from both parents are
68 activated. It has been proposed that bi-parental TE rerepression results from the accumulation
69 of genetic incompatibilities in piRNA pathway proteins over evolutionary time, which impairs TE
70 silencing mechanisms in interspecific hybrids (Kelleher et al. 2012). Such bi-parental TE
71 derepression has been reported in Drosophila interspecific hybrids (Labrador et al. 1999;
72 Kelleher et al. 2012; Romero-Soriano et al. 2017), as well as in some mammal, plant and fish
73 hybrids (O'Neill et al. 1998; Ungerer et al. 2006; Dion-Côté et al. 2014). While TE derepression
74 has been observed in various aforementioned hybrid systems, its downstream genomic impact
75 remains unclear. In some cases, studies have focused solely on transcriptional derepression,
76 leaving open the question of whether increased TE expression leads to greater TE
77 accumulation or transposition. Moreover, many studies rely on later-generation hybrids, making
78 it difficult to determine whether derepression occurred in the initial F_1 or F_2 generations, or
79 emerged later. In other cases, the observed genomic instability may be confounded

80 by chromosomal rearrangements or the effects of meiotic recombination, which cannot be
81 excluded as alternative or contributing mechanisms.

82 *Fundulus heteroclitus* and *F. diaphanus* are killifish species that diverged between 15
83 and 25 million years ago (Ghedotti and Davis 2017). While *F. diaphanus* is primarily found in
84 freshwater, *F. heteroclitus* preferentially occurs in saltwater (Scott and Crossman 1973). Both
85 species are euryhaline and co-occur in brackish water, where sexually and asexually
86 reproducing hybrids have been reported (Fritz and Garside 1974; Dawley 1992; Hernández
87 Chávez and Turgeon 2007; Mérette et al. 2009). An opening formed ~70 years ago (~25-40
88 generations) in Porters Lake, NS, allowing sea water to enter (Fritz and Garside 1974).
89 Therefore, secondary contact between *F. heteroclitus* and *F. diaphanus* in Porters Lake is very
90 recent, from ~70 years to a conservative maximum of 13 000 years, as the area was covered by
91 Wisconsinian glaciers until then (Shaw et al. 2006; Dalziel et al. 2020). There, previous work
92 has revealed the presence of asexually reproducing "frozen F1" hybrids that reproduce by
93 gynogenesis, carrying a full, unrecombined chromosome set from each parental species due to
94 disrupted meiotic recombination between homologs (Hernández Chávez and Turgeon 2007;
95 Dalziel et al. 2020, Dion-Côté et al., *unpublished observations*). Several clonal lineages co-exist,
96 but one lineage is predominant (Hernández Chávez and Turgeon 2007; Dalziel et al. 2020).
97 This system provides a rare opportunity to investigate maternally- and paternally-transmitted TE
98 accumulation in "frozen F1" hybrids, controlling for some of the confounding effects of meiotic
99 recombination and large-scale chromosome rearrangements. The relatively young age of the
100 *Fundulus* clones also allows to capture TE accumulation in these hybrids, if any, before their
101 decay, as predicted under long-term asexual reproduction (Hickey 1982; Arkhipova and
102 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 microsatellite 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

124 (v0.38) using the following parameters: ILLUMINACLIP 2:30:7:4, HEADCROP: 10 and MINLEN:
125 100.

126 *Repeat characterization*

127 We characterized the most abundant repeats in hybrid individuals in a single run of
128 RepeatExplorer2 (Novák et al. 2013). We randomly selected and concatenated 71,429 trimmed
129 read pairs from each of the seven hybrid individuals for a total of 500,000 read pairs. These
130 reads were grouped by RepeatExplorer2 into clusters. We further classified each cluster into the
131 most specific group possible, based on a hierarchical decision tree (Fig. S1). For TEs, a cluster
132 is equivalent to a complete or partial subfamily. “Unclassified repeats” are clusters that were not
133 annotated as a satellites, mobile elements or rDNA, whereas “Unknown” clusters were not
134 classified as organelles or contamination. The final repeat database was composed of 14,011
135 contigs (Supplementary File 1).

136 We used the “id_rmasker.py” script to add the annotation to the sequence IDs in
137 RepeatMasker’s format (Smit et al. 2013-2015) and the “replace_patterns.py” script to transform
138 the RepeatExplorer2 annotation to be cluster-specific (https://github.com/fjruizruano/ngs-
139 protocols). We used the resulting repeat database to estimate the proportion of the genome
140 occupied by each cluster. To do this, we selected 5 million trimmed read pairs per sequenced
141 library using seqtk v1.3 (https://github.com/lh3/seqtk). We then quantified the abundance of
142 each repeat within the database with this subset of reads using RepeatMasker v4.1.3
143 (Supplementary Files 2–3). RepeatMasker was run using the repeat_masker_run_big.py script
144 from the satMiner toolkit (https://github.com/fjruizruano/satminer, last access on 11/07/2022)
145 (Ruiz-Ruano et al. 2016). This script runs RepeatMasker on large query files by splitting them
146 up into subsets. It also runs the calcDivergenceFromAlign.pl script from RepeatMasker, which

147 counts the total number of nucleotides occupied by each group of repeats along divergence
148 intervals (Kimura substitution levels) of 1% needed to generate repeat landscapes. Genome
149 proportion occupied by each cluster was obtained by normalizing the number of nucleotides
150 occupied by each cluster to the total number of nucleotides for its subset of reads. We then
151 generated average repeat landscapes for each group (*F. heteroclitus*, *F. diaphanus*, and
152 hybrids) and subtractive repeat landscapes as the average values of a group minus the average
153 values of the other group. A principal component analysis was run on the abundance for each
154 repeat category using the function PCA() from the FactoMineR package (Lê et al. 2008) in R.
155 We calculated the average genome proportion of the parental species and the hybrids for each
156 cluster. We then detected important deviations from the observed genome proportion occupied
157 by each repeat cluster in hybrids compared with the expected proportion as the average
158 between the genome proportion of both parental species. This analysis revealed repeat.CL27
159 as being ~4X more abundant than expected.

160 *Haplotype calling*

161 We tested whether some species-specific subfamilies of repeat.CL27 preferentially
162 accumulated in hybrids using the sequence of ORF2 (reverse transcriptase, see *Copy number*
163 *validation by qPCR*) from the repeat.CL27 consensus sequence as a reference. We mapped the
164 trimmed paired reads from the 14 individuals using bwa-mem2 v.2.2.1 with relaxed conditions
165 ("-SP -k 14 -L 0,0 -U 0 -B 1"). We processed the resulting BAM files with samtools v1.14
166 (Danecek et al. 2021) as follows. First, we sorted and kept reads mappings with MAPQ ≥ 40
167 and merged the mapped reads for each parental species separately. We then added the read
168 group for each BAM file with Picard v3.1.1 (Broad Institute). We performed haplotype calling
169 with freebayes v1.3.2 using a minimum length of 7 (--haplotype-length 7) (Garrison and Marth
170 2012). We extracted haplotypes from the resulting VCF file with BCFtools query v.1.10.2

171 (Danecek et al. 2021) applying the option (-f
172 '%CHROM\t%POS\t%REF\t%ALT\t[%SAMPLE:%GT\t]\n'). Then, we extracted private
173 haplotypes from each parental species and selected haplotypes longer than 15 nt. Finally, we
174 extracted the mapped reads from the 14 sequencing libraries (parental species and hybrids) in
175 FASTQ format with samtools fastq and counted the number of occurrences of each haplotype
176 with grep (Supplementary File 4).

177 *Copy number validation by qPCR*

178 We identified open reading frames (ORF) in the consensus sequence of repeat.CL27
179 (CL27Contig27_13_sc_0.609135_l_5617) using ORFfinder
180 (https://www.ncbi.nlm.nih.gov/orffinder/), which identified two long non-overlapping ORFs
181 (ORF1 and ORF2, 420 and 2004 nt, respectively). We then estimated the number of copies in
182 18 individuals (six individuals for each group) by qPCR. We designed primers targeting the
183 ORF2 from repeat.CL27 using Primer-BLAST (F: 5'-AGGAACCTGAAAAGTGGAGCA-3', R: 5'-
184 TGGTGCCGAATTGAGAGAGG3') and the citrate synthase gene from the *F. heteroclitus*
185 reference genome (NC_046361, F: 5'-TGCACCCCATGTCTCAGTTC-3', R: 5'-
186 TTGGACTTGTGCACTCCCTC-3).

187 Efficiency curves were generated using 0.0032 - 10 ng (5-fold serial dilution) of gDNA
188 from a single *F. heteroclitus* female individual. Reactions (7 µl) contained 0.25 µM of each
189 primer and 1X Luna Universal qPCR master mix (New England BioLabs). Amplification began
190 with an initial denaturation at 95°C for 1 min, followed by denaturation at 95°C for 15 sec and
191 combined annealing and amplification at 60°C for 30 sec, for a total of 40 cycles. Amplification
192 efficiency was acceptable for both genes (97.3%, R^2 = 0.994 for citrate synthase, 100.65%, R^2 =

193 0.994 for ORF2 of repeat.CL27) and melting curves revealed a single peak per reaction. Both
194 reactions were run on the same plate for each sample.

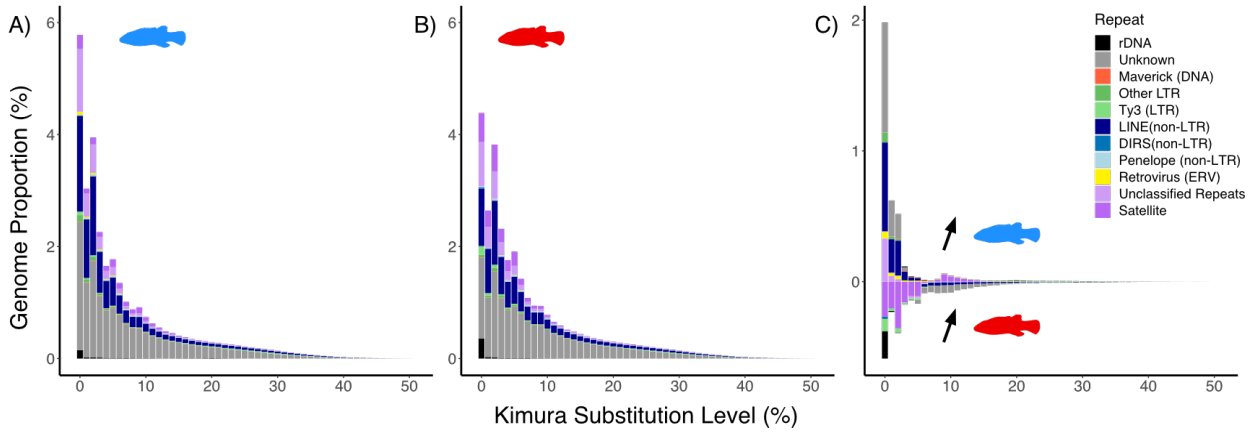
195 *Cloning and sequencing of repeat.CL27 ORF2*

196 We manually designed primers to amplify the full-length reverse transcriptase encoded
197 by ORF2 from the consensus sequence of repeat.CL27 (F: 5'-
198 ATGAGGTATGACCTGCAAAAATACC-3', R: 5'-GATCCAAATCCTTTCAGCCCTTC-3'). The
199 reactions (25 μl) contained 0.6 μM of each primer, 2 mM $MgCl_2$, 2 mM dNTPs, 1.25 U GoTaq
200 and 45 ng of DNA. The amplification protocol included an initial denaturation step at 95°C for 2
201 min, followed by 32 cycles of denaturation at 95°C for 1 min, annealing at 60°C for 1 min and
202 amplification at 72°C for 1 min, with a final elongation at 72°C for 5 min. Amplicons were gel-
203 purified using the Monarch DNA gel extraction kit (New England BioLabs). Purified amplicons
204 were cloned using the TOPO TA cloning kit for sequencing (Invitrogen) following the
205 manufacturer's recommendations and transformed in DH5α competent cells (Fisher Scientific).
206 Sequencing of the plasmids was performed by Plasmidsaurus, and vector sequence was
207 removed using A Plasmid Editor (Davis and Jorgensen 2022) (Supplementary file 5).

208 **Results and Discussion**

209 *F. diaphanus* had a slightly higher repetitive DNA load than *F. heteroclitus* (30.39% vs
210 29.30%, $p < 0.001$, Type III ANOVA, Table S4), consistent with its larger genome (1.48 Gb for
211 *F. diaphanus* vs 1.33 Gb for *F. heteroclitus*, Dawley 1992). For *F. heteroclitus*, the proportion of
212 the genome occupied by TEs was lower than previously reported using an assembly-based
213 approach, which is more sensitive to low copy number repeats in the assembled genome
214 (32.66%, Reinar et al. 2023). In fishes, TEs are highly diversified, with non-LTR
215 retrotransposons and DNA transposons typically being the most abundant, with many families

216 that remain active (Volff 2005; Shao et al. 2019). As expected, non-LTR retrotransposons were
217 the most abundant TEs in both species. On the contrary, we identified very few DNA
218 transposons (<1% of the genome for Maverick elements and no hATs), but this would be
219 expected if many of them were non-autonomous, i.e., lacking any ORFs (Peona et al. 2024).
220 We compared the repetitive DNA content of *F. diaphanus* and *F. heteroclitus* by generating
221 repeat landscapes which revealed higher abundances among the lower Kimura substitution
222 level bins for both species, consistent with a recent accumulation of repeats (Fig. 1A & B). A
223 subtractive repeat landscape, calculated as *F. diaphanus* minus *F. heteroclitus*, revealed that
224 the main differences between both species were found among the less divergent repeats, i.e.,
225 those that expanded more recently (Fig. 1C). This analysis also revealed an expansion of many
226 TE groups as well as unknown/unclassified repeats in *F. diaphanus* (or a contraction in *F.*
227 *heteroclitus*) whereas young satellite DNA appeared to have recently expanded in *F.*
228 *heteroclitus* (or contracted in *F. diaphanus*).



229

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

237 To visualize how repeat content in hybrids differs from the parental species, we
238 performed a principal component analysis (PCA) on the genome proportion occupied by each
239 repeat group. Principal components (PC) 1 (66.1% of the variance) clearly separated parental
240 species, which was expected given their contrasted repeat landscapes (Fig. 2A). Along PC1,
241 hybrids were also intermediate between parental species, although slightly closer to *F.*
242 *diaphanus*. This was expected as *F. diaphanus* has a larger genome. PC2 (25.4% of the
243 variance) clearly separates hybrids from parental species, with all repeat groups being positively
244 correlated to this PC, except for rDNA (more abundant in *F. heteroclitus*) and retroviruses (more
245 abundant in *F. diaphanus*). This suggests that most repeat groups accumulated in hybrid
246 genomes. We thus generated a subtractive repeat landscape to compare the observed hybrid
247 repeat content with the "expected" repeat content (i.e., the sum of half the average abundance
248 of each parental species for each repeat group, Fig. 2B). This revealed that hybrids had a
249 higher repetitive DNA content than expected for every repeat group, except rDNA (Fig. 2C). Yet,
250 the genome proportion occupied by repeats in hybrids was only 0.665% higher than in *F.*
251 *diaphanus* and 1.69% in *F. heteroclitus* (Table S4). These results suggest that most repetitive
252 elements have accumulated in *Fundulus* hybrids, consistent with suboptimal piRNA pathway
253 protein interaction for TEs, although very moderately. While we cannot entirely exclude
254 technical biases, such as underestimating hybrid genome size or overestimating repeat
255 abundance through parental allele separation, this pattern would not be restricted to the
256 youngest repeats (i.e., lowest Kimura substitution bins) as we observe. Another caveat is that
257 the individuals from the parental species that we sequenced may not fully capture the variation
258 present in the parental individuals that produced the sequenced clones.

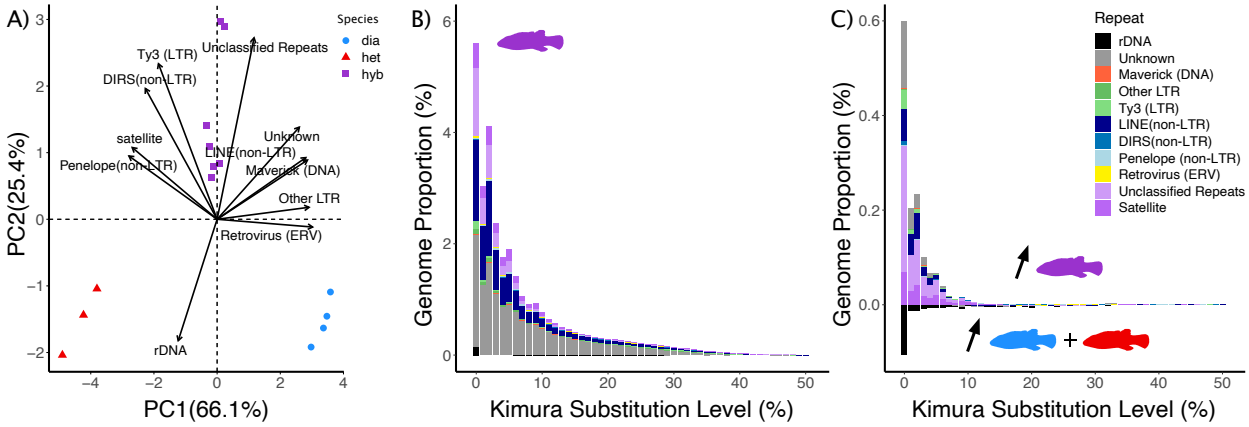


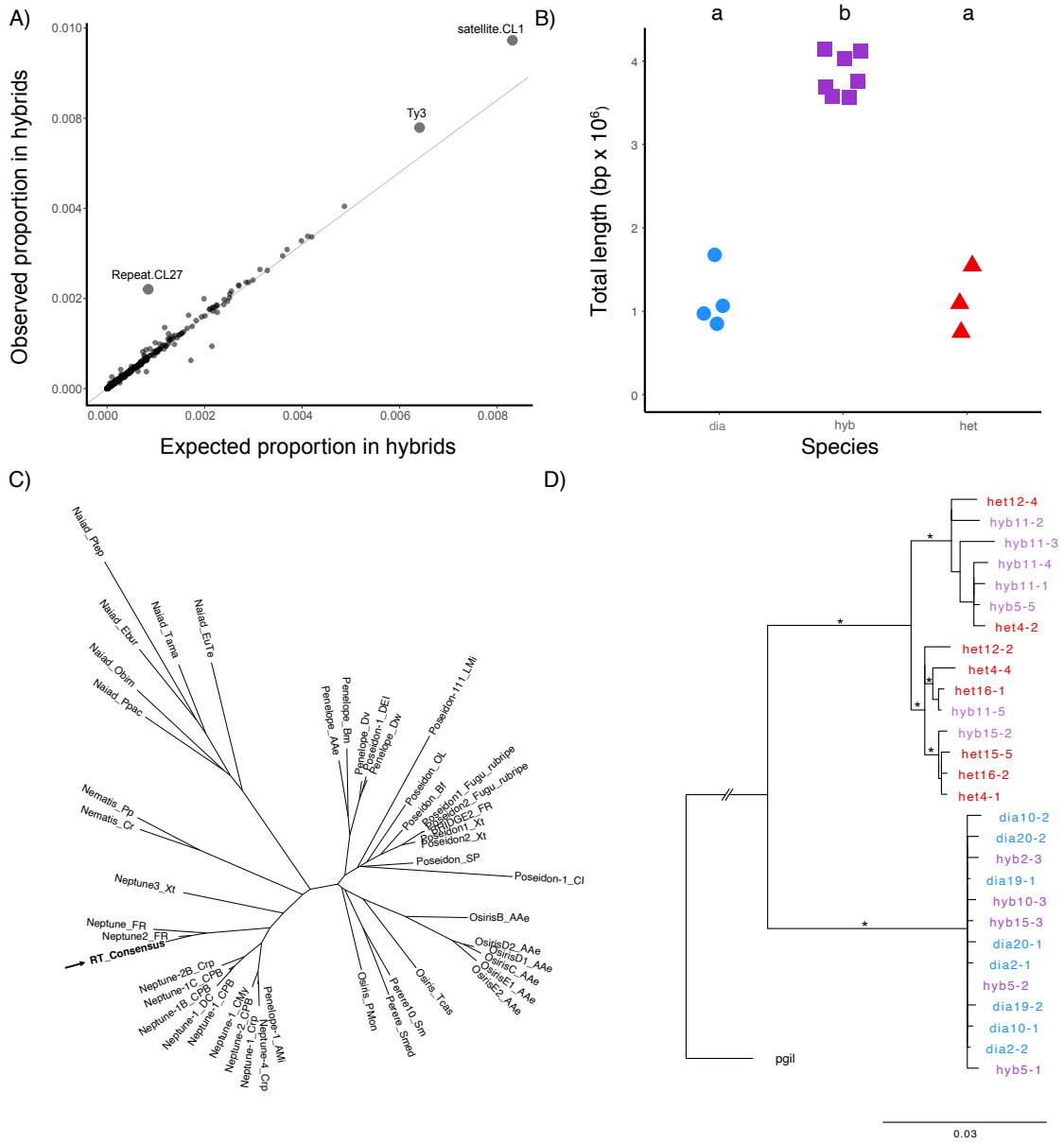
Figure 2. Repeat content in *Fundulus* hybrids. A) PCA of the repeat group content on *Fundulus* fish. B) Repeat landscape of *Fundulus* hybrids. C) Subtractive repeat landscape generated by subtracting expected abundance values (half the average abundance in *F. diaphanus* plus half the average abundance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

Therefore, we examined whether some repeats expanded more extensively in hybrids, compared to our expectation (i.e., the sum of half the average abundance of each parental species for each repeat group, Fig. 3A). We extracted the only three repeat clusters that varied by more than 1 Mb between the observed and expected hybrid repeat content (Fig. 3A). The first one, annotated as a Ty3 element (LTR retrotransposon), was ~1.15 times more abundant in hybrids than expected, indicating a moderate increase. Although satellite.CL1 was more abundant in hybrids than expected (observed: $1.37 \times 10^7 \pm 1.13 \times 10^6$ bp, expected: 1.17×10^7 bp), it fell within the *F. heteroclitus* range ($1.35 \times 10^7 \pm 1.15 \times 10^6$ bp). A third candidate, repeat.CL27 ("unclassified", Fig. S1), stood out, being ~4 times more abundant than expected in hybrids (Fig. 3B, $p < 0.001$, Type III ANOVA,). This result was further validated by qPCR with different individuals, to exclude methodological biases, which revealed a 2.92 average increased abundance in hybrids compared to parental species (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*) (subfamily Neptune-2_GA, positions 5128-2460, 66.03% identity). Elements from the

280 Neptune superfamily are poorly characterized *Penelope*-like elements defined by the presence
281 of a reverse transcriptase (RT) and a GIY-YIG endonuclease domain (Evgen'ev and Arkhipova
282 2005) that have now been detected across algae, plants, fungi, protists and animals (Craig et al.
283 2021). We identified two long ORFs in repeat.CL27, from position 73-492 (419 nt, ORF1) and
284 from position 666-2669 (2003 nt, ORF2). We did not detect sequence similarity for ORF1 in any
285 of the queried databases (Repbase, GenBank, gtRNAdb, CDD), but ORF2 contains the intact
286 RT (AA 590-558) and GIY-YIG (AA 299-392) domains based on the Conserved Domains
287 Database (CDD). A phylogeny with other PLEs revealed that ORF2 groups with other Neptune
288 subfamilies (Fig. 3C). We cloned and sequenced full ORF2 from various individuals, which
289 revealed the existence of two subfamilies of this Neptune element that are specific to each
290 parental species (Fig. 3D). Both subfamilies were found in hybrids, which prompted us to
291 examine the accumulation of short, species-specific haplotypes in repeat.CL27 from the
292 sequencing reads (Supplementary File 4). We detected 8 regions > 15 nt, resulting in 27
293 haplotypes, 12 of which were specific to *F. diaphanus* and 7 to *F. heteroclitus* (i.e., short
294 haplotypes that are diagnostic of a subfamily that is restricted to either parental species).
295 Among *F. diaphanus*-specific haplotypes, five out of seven were more abundant in hybrids than
296 expected (observed abundance/expected abundance median = 6.79). Likewise, five out of 12 *F.*
297 *heteroclitus*-specific haplotypes showed a higher abundance in hybrids than expected
298 (observed abundance/expected abundance median = 7.12). Thus, some species-specific
299 haplotypes are present in excess in these asexually reproducing "frozen F1" hybrids, whereas
300 others are under-represented. Species-specific haplotypes may be under-represented in hybrid
301 genomes because they were absent in the original parents, emerged in parental species after
302 hybrid formation, or were lost in hybrids through mutations or deletions. Altogether, these results
303 suggest that species-specific variants of this Neptune subfamily from each parental species are

304 accumulating in *Fundulus* hybrids. We propose that this accumulation may be restricted to

305 currently active variants.



306

307 **Figure 3.** A Neptune transposable element accumulates in *Fundulus* hybrids. A) Observed
308 proportion of the genome occupied by each repeat cluster in hybrids as a function of the
309 expected copy number. Repeat.CL27 is highlighted. The line depicts a 1:1 ratio. B) Total
310 genomic length (bp) occupied by repeat.CL27 based on sequencing ($p < 0.001$, Type III
311 ANOVA). C) Phylogeny of PLEs RT and repeat.CL27 ORF2 (RT consensus). PLEs sequences
312 were graciously provided by Tomás Carrasco-Valenzuela (Carrasco-Valenzuela et al. 2023). D)

313 Phylogeny of cloned full-length RT from 14 individuals. Outgroup sequence from *Poecilia gillii*
314 (NCBI accession number CAEGAJ010041810.1: 443 - 2446) (van Kruistum et al. 2021) was
315 obtained by doing a BLASTN search of the consensus sequence against the WGS database
316 restricted to the *Poecilia* genus (taxid:8080). Sequences were aligned using MAFFT-LINSI
317 v7.505 (Katoh and Standley 2013) (Supplementary file 6). We manually inspected the resulting
318 alignment for the presence of PCR chimeras and removed them. We built a phylogeny using IQ-
319 TREE v2.4.0 with default parameters and 1000 bootstraps (Nguyen et al. 2015). Trees were
320 edited using Figtree v1.4.4 (https://github.com/rambaut/figtree). Branches with >0.85 bootstrap
321 support are indicated with an asterisk.

322 **Conclusion**

323 While the moderate accumulation of many repetitive sequences we observed in hybrids
324 is coherent with suboptimal piRNA pathway protein interactions (Kelleher et al. 2012), technical
325 biases cannot be entirely excluded. We identified two Neptune subfamilies, each of which are
326 specific to either *F. diaphanus* or *F. heteroclitus*, that accumulated dramatically in naturally
327 occurring "frozen F1" hybrids. This pronounced accumulation of two TE subfamilies, each
328 inherited from either parental species, differs from previously described patterns:
329 in *Drosophila* dysgenic crosses, where only paternally inherited TEs are derepressed in the
330 progeny of naïve females, and in other interspecific hybrids, where multiple TE families tend to
331 accumulate (e.g., Kidwell et al. 1977; Labrador et al. 1999; Kelleher et al. 2012). The strong
332 accumulation of this Neptune element suggests that this element is particularly aggressive.
333 Because naturally occurring sexually and asexually reproducing hybrid lineages have
334 repeatedly evolved and can be compared with new lineages produced in the lab, we further
335 propose that the *Fundulus* complex is an excellent system in which to disentangle hybridization-
336 driven and asexuality-mediated effects on TE dynamics in the future.

337

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.

387 Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic
388 shock involving transposable elements reactivation in hybrids of young lake whitefish
389 species. *Mol. Biol. Evol.* 31:1188–1199.

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

392 Evgen'ev MB, Arkhipova IR. 2005. Penelope-like elements--a new class of retroelements:
393 distribution, function and possible evolutionary significance. *Cytogenet. Genome Res.*
394 110:510–521.

395 Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, Corces VG.
396 1997. Penelope, a new family of transposable elements and its possible role in hybrid
397 dysgenesis in Drosophila virilis. *Proc. Natl. Acad. Sci. U. S. A.* 94:196–201.

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

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

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

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

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

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

412 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
413 improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

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

416 Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid dysgenesis in *Drosophila melanogaster*: A
417 syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics*
418 86:813–833.

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

421 van Kruistum H, Nijland R, Reznick DN, Groenen MAM, Megens H-J, Pollux BJA. 2021. Parallel
422 genomic changes drive repeated evolution of placentas in live-bearing fish. *Mol. Biol. Evol.*
423 38:2627–2638.

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

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

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

431 Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ.
432 2009. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila
433 ovary. *Cell* 137:522–535.

434 Marie, Pauline P., Stéphane Ronsseray, and Antoine Boivin. 2017. From Embryo to Adult:
435 piRNA-Mediated Silencing throughout Germline Development in Drosophila. G3 (Bethesda,
436 Md.) 7 (2): 505–16.

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

440 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective
441 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*
442 32:268–274.

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

446 O'Neill RJ, O'Neill MJ, Graves JA. 1998. Undermethylation associated with retroelement
447 activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature*
448 393:68–72.

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

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

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

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

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

462 Ryazansky, Sergei, Elizaveta Radion, Anastasia Mironova, Natalia Akulenko, Yuri Abramov,
463 Valeriya Morgunova, Maria Y. Kordyukova, Ivan Olovnikov, and Alla Kalmykova. 2017.
464 Natural Variation of piRNA Expression Affects Immunity to Transposable Elements. PLoS
465 Genetics 13 (4): e1006731.

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

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

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

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

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

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

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