Genome of the Rio Pearlfish (*Nematolebias whitei*), a bi-annual killifish model for Eco-Evo-Devo in extreme environments

Andrew W. Thompson (iD) ,[1,2,*] Harrison Wojtas,[1] Myles Davoll,[1,3] and Ingo Braasch (iD) [1,2,*]

[1]Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA,
[2]Ecology, Evolution & Behavior (EEB) Program, Michigan State University, East Lansing, MI 48824, USA,
[3]Department of Biology, University of Virginia, Charlottesville, VA 22903, USA

*Corresponding authors: Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA. Email: thom1524@msu.edu (A.W.T.); Department of Integrative Biology, Michigan State University, East Lansing, MI 48824, USA. Email: braasch@msu.edu (I.B.)

Abstract

The Rio Pearlfish, *Nematolebias whitei*, is a bi-annual killifish species inhabiting seasonal pools in the Rio de Janeiro region of Brazil that dry twice per year. Embryos enter dormant diapause stages in the soil, waiting for the inundation of the habitat which triggers hatching and commencement of a new life cycle. Rio Pearlfish represents a convergent, independent origin of annualism from other emerging killifish model species. While some transcriptomic datasets are available for Rio Pearlfish, thus far, a sequenced genome has been unavailable. Here, we present a high quality, 1.2 Gb chromosome-level genome assembly, genome annotations, and a comparative genomic investigation of the Rio Pearlfish as representative of a vertebrate clade that evolved environmentally cued hatching. We show conservation of 3D genome structure across teleost fish evolution, developmental stages, tissues, and cell types. Our analysis of mobile DNA shows that Rio Pearlfish, like other annual killifishes, possesses an expanded transposable element profile with implications for rapid aging and adaptation to harsh conditions. We use the Rio Pearlfish genome to identify its hatching enzyme gene repertoire and the location of the hatching gland, a key first step in understanding the developmental genetic control of hatching. The Rio Pearlfish genome expands the comparative genomic toolkit available to study convergent origins of seasonal life histories, diapause, and rapid aging phenotypes. We present the first set of genomic resources for this emerging model organism, critical for future functional genetic, and multiomic explorations of "Eco-Evo-Devo" phenotypes of resilience and adaptation to extreme environments.

Keywords: *Nematolebias whitei*; Rio Pearlfish; diapause; aging; hatching; extreme environments; Eco-Evo-Devo; teleost

Introduction

Aplocheiloid killifishes inhabit tropical freshwater habitats around the world. Some African and Neotropical species live in ephemeral waters that are subject to seasonal desiccation (Myers 1952; Simpson 1979). Desiccation kills the adults, but embryos survive inside specialized eggs (Thompson *et al.* 2017a) buried in the soil via 3 diapause stages (DI, DII, DIII; Wourms 1972a,1972b, 1972c). DI occurs as a migratory dispersion of blastomeres, DII occurs during somitogenesis when organs are rudimentary, and DIII occurs after organogenesis when the embryo is fully formed and poised to hatch upon habitat inundation. This seasonal life history is a remarkable example of convergent evolution (Furness *et al.* 2015) with 7 gains across killifish evolution (Thompson *et al.* 2021a). Additionally, annual killifishes show rapid aging due to relaxed selection on lifespan (Cui *et al.* 2019) and are an important emerging model system for the study of senescence (Valenzano *et al.* 2011, 2015; Harel *et al.* 2015; Reichwald *et al.* 2015).

The Rio Pearlfish, *Nematolebias whitei*, is a seasonal killifish endemic to the coastal plains of the Rio de Janeiro region in Brazil, inhabiting pools that dry twice annually, from July–August and February–March (Fig. 1a; Myers 1942; Costa 2002). Pearlfish represents a separate origin of seasonality from other killifish model species, i.e. *Nothobranchius furzeri* and *Austrofundulus limnaeus* (Thompson *et al.* 2021a). In *N. whitei*, DI and DII are facultative, and DIII is a "prolonged," "deep" stasis compared to hatching delay and DIII in other killifishes, occurring just before environmentally cued hatching upon submersion in water (Wourms 1972c; Thompson and Ortí 2016; Thompson *et al.* 2017b). Pearlfish was suggested as a top candidate species for killifish models in the seminal work of developmental biologist John P. Wourms in 1967. They are small, prolific, and hardy, and spawn in sand (Wourms 1967), making them easily reared laboratory animals that are furthermore amenable to genetic manipulation like other killifishes (Aluru *et al.* 2015; Harel *et al.* 2015). Pearlfish has also been an emergent system to study aging (Ruijter 1987), environmental
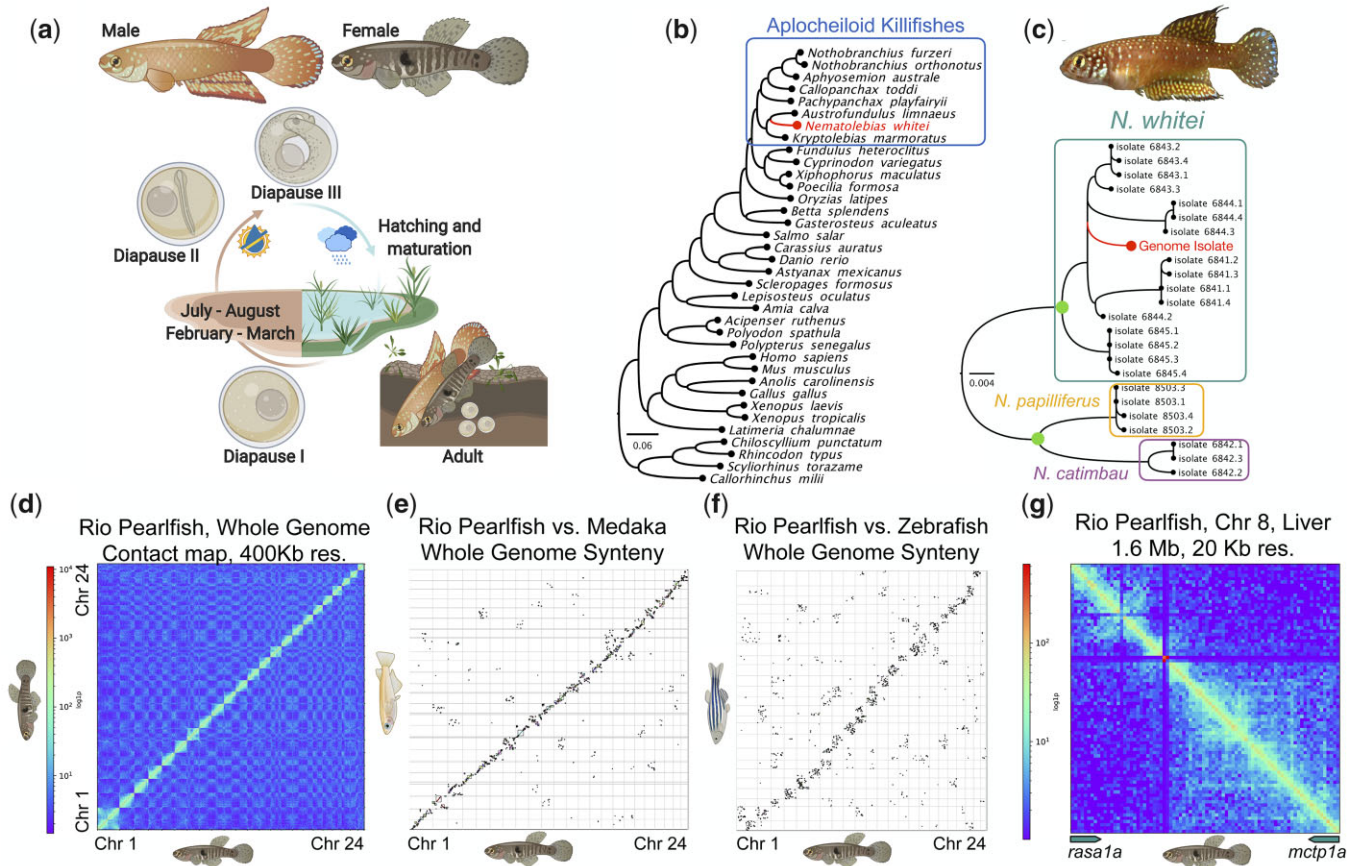
Fig. 1. Rio Pearlfish evolution, ecology, development, and 3D genome structure. a) Bi-annual life cycle of the Rio Pearlfish with three developmental diapause stages following burying of eggs in soil. b) Relative position of Rio Pearlfish in the vertebrate tree of life inferred by Orthofinder based on annotated proteins. c) DNA barcode (*cox1* and *cytb*) phylogeny inferred with RAxML of the genus *Nematolebias* confirming the identity of the genome specimen as *N. whitei*. Sequences from Costa *et al.* (2014) were used for comparison to the reference genome sequence. Green nodes show 100% bootstrap support for the reciprocal monophyly of *N. whitei* with other genera and confirms the identity of the genome specimen with high confidence. d) Hi-C contact map of the Rio Pearlfish genome showing linkage of the 24 chromosomes into chromosomal pseudomolecules. e, f) SynMap genome-wide synteny plots of Rio Pearlfish *vs.* medaka (e) and *vs.* zebrafish (f) showing genome-structure conservation across over 250 million years of teleost evolution. g) Hi-C contact maps of the syntenic region between *rasa1a* and *mctp1a* in Pearlfish liver tissue. These contact maps highlight the conserved 3D structure that include TADs conserved across teleost evolution as well as cell types and developmental stages (Nakamura *et al.* 2021). Species graphics Created with BioRender.com.

influences on development (Ruijter *et al.* 1984), the role of prolactin in hatching control (Schoots *et al.* 1983; Ruijter and Creuwels 1988), resilience to perturbations in development with the ability to develop normally from diblastomeric eggs (Carter and Wourms 1993), and the transcriptional control of diapause and hatching (Thompson and Ortí 2016).

Here, we construct a chromosome-level genome assembly for the Rio Pearlfish, utilizing Hi-C contact maps, genome annotations, and gene expression analyses to characterize genomic evolution and hatching biology in this extremophile vertebrate.

Materials and methods
Genome sequencing and assembly

All animal work was approved by the Michigan State University Institutional Animal Care and Use Committee (PROTO202000108).

A total of 1.25 ng of template genomic DNA extracted from the liver of a single adult female *N. whitei* was loaded on a Chromium Genome Chip. Whole-genome sequencing libraries were prepared using 10× Genomics Chromium Genome Library & Gel Bead Kit v.2, Chromium Genome Chip Kit v.2, Chromium i7 Multiplex Kit, and Chromium controller according to the manufacturer's instructions with one modification. Briefly, gDNA was combined with Master Mix, a library of Genome Gel Beads, and partitioning oil to create Gel Bead-in-Emulsions (GEMs) on a Chromium Genome Chip. The GEMs were isothermally amplified. Prior to Illumina library construction, the GEM amplification product was sheared on a Covaris E220 Focused Ultrasonicator to ∼350 bp and then converted to a sequencing library following the 10× standard operating procedure. A total of 679.43 million read pairs were sequenced on an Illumina HiSeqX sequencer, and a *de novo* assembly was constructed with Supernova 2.1.1 (Weisenfeld *et al.* 2018).

A Chicago library was prepared as described previously (Putnam *et al.* 2016). Briefly, ∼500 ng of HMW gDNA was reconstituted into chromatin *in vitro* and fixed with formaldehyde. Fixed chromatin was digested with DpnII, the 5′ overhangs were filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed, and the DNA was purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then

sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeqX to produce 242 million 2× 150 bp paired end reads.

A Dovetail Hi-C library was prepared in a similar manner as described previously (Lieberman-Aiden *et al.* 2009). For each library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DpnII, the 5′ overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed, and the DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeqX to produce 179 million 2 × 150 bp paired end reads.

The Supernova de novo assembly built from 10× Chromium data, Chicago library reads, and Dovetail Hi-C library reads were used as input data for assembly scaffolding with HiRise v1 (Putnam *et al.* 2016). An iterative analysis was conducted. First, Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper (http://snap.cs.berkeley.edu). The separations of Chicago read pairs mapped within draft scaffolds were analyzed by HiRise v1 to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold. After aligning and scaffolding Chicago data, Dovetail Hi-C library sequences were aligned, and scaffolds were generated following the same approach.

Genome annotation

The Rio Pearlfish genome was annotated with the NCBI Euakryotic genome annotation pipeline v9.0 (Thibaud-Nissen *et al.* 2016) and with MAKER 2.31 (Cantarel *et al.* 2008; Campbell *et al.* 2014; Bowman *et al.* 2017) using protein evidence from 15 fish species (Supplementary Table 1) and transcriptome evidence from Rio Pearlfish DIII embryos and hatched larvae (Thompson and Ortí 2016). Genome assembly and annotation completeness (Supplementary Table 2) were analyzed with BUSCO v5 (Simão *et al.* 2015) and CEGMA 2.4 (Parra *et al.* 2007) via the gVolante server (Nishimura *et al.* 2017, https://gvolante.riken.jp).

Phylogenetics and orthology

To confirm species identification, we extracted and concatenated the barcoding marker genes *cox1* and *cytb* from our genome assembly, aligned them with orthologous sequences from all 3 described *Nematolebias* species (Costa *et al.* 2014) and inferred a phylogeny partitioned by codon and gene with RAxML (Stamatakis 2006, 2014) with the following parameters: -T 4 -N autoMRE -m GTRCAT -c 25 -p 12345 -f a -x 12345 –asc-corr lewis. We used Orthofinder v2.4.1 (Emms and Kelly 2015) to identify orthologous protein sequences between *N. whitei* and 35 other vertebrates genomes (Supplementary Table 3) as well as protein sequences obtained from Cui *et al.* (2019), Hara *et al.* (2018), and the longest isoforms of other species available on NCBI RefSeq (last accessed 2021 September 22) downloaded with ortholog's retrieve longest isoforms function (Drost *et al.* 2015). The output of

Orthofinder (Supplementary Table 4) was examined to identify Pearlfish-specific orthogroups. Genes in these orthogroups were used as queries in BLAST searches (e value cutoff of e−3) against Japanese medaka (HdrR strain, assembly ASM223467v1) protein sequences downloaded from Ensembl (last accessed 2022 January 17, Supplementary Table 5). We performed a statistical overrepresentation test with a Fisher's exact test and a false discovery rate correction on the Gene Ontologies (GOs) of these medaka genes using Panther v.16.0 (Mi *et al.* 2021) with the GO biological processes complete database (Supplementary Table 6).

Synteny and genome 3D structure

We examined conservation of synteny using genome assemblies and NCBI annotations for Rio Pearlfish, medaka (oryLat2, UCSC), and zebrafish (GCF_000002035.5_GRCz10, NCBI) as input for SynMap in the online CoGe database and toolkit (Lyons and Freeling 2008; last accessed October 14, 2021). Bwa v 0.7.17 (Li and Durbin 2009) was used to independently map Rio Pearlfish Hi-C read pairs to the genome assembly with the following parameters: bwa mem -A1 -B4 -E50 -L0, and HiCExplorer 3.6 (Ramírez *et al.* 2018) was used to construct a Hi-C matrix with the resulting bam files as follows: hicBuildMatrix –binSize 10000 –restrictionSequence GATC –danglingSequence GATC. The matrix was corrected via hicCorrectMatrix correct –filterThreshold -1.5 5. The matrix was binned depending on preferred resolution for viewing. Contact maps were visualized with hicPlotMatrix –log1p, and compared with contact maps of syntenic regions in medaka and zebrafish (Nakamura *et al.* 2021).

Repeat content and repeat landscape

We constructed a species-specific repeat database with Repeat Modeler 2.0.1 (Smit and Hubley 2008). This library as well as vertebrate Repbase annotations (Jurka 2000) (downloaded 15 November 2017), and repeat libraries from platyfish (Schartl *et al.* 2013), coelacanth (Amemiya *et al.* 2013), bowfin (Thompson *et al.* 2021b), and spotted gar (Braasch *et al.* 2016) were combined to annotate repeat elements with RepeatMasker v4.0.5 (Smit *et al.* 2013). CalcDivergenceFromAlign.pl and createRepeatLandscape.pl in the RepeatMasker package were used to generate a repeat landscape. We graphically compared the repeat landscape of Rio Pearlfish to those described for other sequenced killifish species (Reichwald *et al.* 2015; Valenzano *et al.* 2015; Rhee *et al.* 2017; Cui *et al.* 2019) to identify similarities and difference in the magnitude and location of peaks at different Kimura distances in the histograms.

Hatching enzyme gene expression

Aquatic vertebrates hatch by secreting choriolytic enzymes from hatching gland cells (HGCs) that break down the egg chorion (Yamagami 1988; Hong and Saint-Jeannet 2014). Teleost fishes underwent hatching enzyme gene duplications followed by divergence and functional divergence into the high choriolytic enzyme (*hce*) and low choriolytic enzyme (*lce*) genes (Yasumasu *et al.* 1992; Kawaguchi *et al.* 2006, 2010; Sano *et al.* 2014). We used BLAST to search the well-studied medaka hatching enzyme paralogs (*lce* and *hce*) against the annotated Pearlfish genome. We used the Pearlfish gene sequences from these BLAST hits as well as other metalloprotease gene sequences from medaka, *Austrofundulus, Kryptolebias*, and *Nothobranchius* (accession numbers in Supplementary Table 7) to infer gene trees. The identified Pearlfish *lce* and *hce* genes are orthologous to those of other teleosts (data not shown). Pearlfish hatching enzyme orthologs were examined for transcript evidence from DIII embryos (Thompson

and Ortí 2016) to identify active *lce* and *hce* gene expression in Pearlfish DIII.

We generated an antisense RNA probe for the Pearfish *lce.2* gene and performed whole-mount RNA *in situ* hybridization to identify hatching enzyme gene expression patterns as markers for the location of HGCs in Pearlfish. Total RNA was extracted from DIII Pearlfish embryos with a Qiagen RNeasy mini plus kit and reverse transcribed with a superscript IV VILO kit (ThermoFisher) according to the manufacturers' instructions. The *lce* cDNA was amplified from the reverse transcribed template via PCR (Primers: Nwh_lce.1_1F: 5′-ATGGACCATAAAG CAAAAGTTTCTCTC-3′; Nwh_lce.1_792R: 5′-CTATTGCTTGTATTT TGAACACTTGT-3′; Nwh_lce.2_1F: 5′-ATGGACCATAAAGCAAAA GTTACTCTT-3′; Nwh_lce.2_825R: 5′-CTATTGCTTGTATTTTGAA CAGTTGT-3′) and *lce.2* was inserted into a TOPO TA cloning kit vector (Invitrogen) according to the manufacturer's instructions. Whole mount *lce.2* mRNA *in situ* hybridization on manually dechorionated DIII Rio Pearlfish embryos was performed following Deyts *et al.* (2005) with a 25 µg/ml proteinase K digestion treatment for 45 min (n = 3 embryos), 60 min (n = 3 embryos), and 90 min (n = 2 embryos).

Results and discussion

Genome sequencing and assembly

We report a high-quality, 1.2 Gb chromosome-level genome assembly of *N. whitei*. The Rio Pearlfish genome assembly consists of 24 chromosomal pseudomolecules represented by 24 superscaffolds that match the described karyotype (n = 24; Van Post, 1965). The assembly has an N50 of over 49.98 Mb and an L50 of 11 scaffolds (Table 1). BUSCO and CEGMA scores for different core gene databases indicate a high-quality assembly with an average of 94% complete BUSCOS and CEGs across all relevant databases (Table 1, Supplementary Table 2).

Genome annotation

The NCBI *N. whitei* Annotation Release 100.20210725 contains 23,038 genes, with 21,341 protein-coding genes, similar to other chromosome-level killifish genome assemblies from *N. furzeri* and *Kryptolebias marmoratus* (Reichwald *et al.* 2015; Valenzano *et al.* 2015; Kelley *et al.* 2016; Rhee *et al.* 2017) (Supplementary Table 8). Minor differences in gene numbers among killifish species are likely due to annotation methods and minor species-specific gene losses or expansions. The number and content of annotated genes can be influenced by evidence used for annotation, differences in gene model prediction likelihoods, and post-annotation filtering (Holt and Yandell 2011; Campbell *et al.* 2014). MAKER annotated 26,016 protein-coding genes, on par with the NCBI annotation. See Table 1 and Supplementary Tables 2 and 8 for Rio Pearlfish genome annotation statistics. Although our BUSCO analyses show fewer genes missed by the NCBI annotation (Supplementary Table 2), our additional MAKER annotation provides additional, valid gene models missed by the NCBI pipeline. For example, MAKER annotates 28 vertebrate and 27 actinopterygian BUSCOs missed by the NCBI annotation pipeline (Supplementary Table 9).

Phylogenetics and orthology

Our orthofinder analysis illustrates the phylogenetic position of Rio Pearlfish among vertebrates (Fig. 1b) and identified 31,317 orthogroups across 36 vertebrate species with 99.2% of Rio Pearlfish genes within orthogroups (Table 1; Supplementary Table 4). We identified 7,287 orthogroups present across all

Table 1. Rio Pearlfish genome assembly (NemWhi1) and annotation statistics.

Genome assembly statistics	
No. of scaffolds	18,999
No. of base pairs	1,218,332,341
N50	49,984,095
L50	11
N90	32,525,398
L90	22
No. of superscaffolds	24
No. of chromosomes (n)[a]	24
GC content	41.8%
Repeat content[b]	57.3%
BUSCO scores[c]: Vertebrata, Actinopterygii	96.9%, 95.5%
CEGMA scores[c]: CEG, CVG	99.19%, 99.57%
Gene annotation statistics	
MAKER annotation	
No. of protein-coding genes	26,016
BUSCO scores[c]: Vertebrata, Actinopterygii	91.4%, 86.2%
NCBI RefSeq annotation[d]	
No. of protein-coding genes	21,341
BUSCO scores[c]: Vertebrata, Actinopterygii	97.4%, 96.5%
No. of genes in orthogroups[e]	21,176 (99.2%)
No. of species-specific orthogroups (genes)	17 (42)

[a] Van Post (1965).
[b] See Supplementary Table 10 for more information.
[c] See Supplementary Table 2 for more information.
[d] See Supplementary Table 8 for more information.
[e] See Supplementary Table 4 for more information.

species from sharks to human to Rio Pearlfish, highlighting the utility of the Rio Pearlfish genome to connect species with extreme developmental phenotypes to other vertebrates, including traditional vertebrate model species such as mouse, *Xenopus*, zebrafish, etc. We confirmed the identity of our genome specimen with barcoding and a molecular phylogeny of *cox1* and *cytb* with its position located within the *N. whitei* clade of *Nematolebias* killifishes (Fig. 1c). We found a total of 17 Pearlfish-specific orthogroups comprising a total of 42 protein sequences. For 39 of these, we established homology to a medaka gene class by BLAST (Supplementary Table 5) and found an overrepresentation for GO terms related to glycolysis (Supplementary Table 6). This may indicate an adaptive expansion of metabolic genes in this species as annual killifishes tolerate anoxia (Podrabsky *et al.* 2007, 2012; Wagner *et al.* 2018), severely depress metabolic rate during diapause (Podrabsky and Hand 1999), and drastically increase metabolic rate during fast maturation (Vrtílek *et al.* 2018) necessary for an annual life cycle. In a separate study, we have also found that genes involved in cell respiration, specifically oxidative phosphorylation, show higher ratios of nonsynonymous/synonymous codon changes in annual killifishes compared to their nonannual counterparts (Thompson *et al.* 2021a). Together, these observations points to potential positive selection on genes involved in cell respiration in annual killifishes.

Synteny and genome 3D structure

Three-dimensional chromatin structure impacts gene regulation and can manifest as topologically associated domains (TADs) that could represent higher order gene regulatory regions conserved across evolution (Krefting *et al.* 2018). However, 3D genome structure has thus far remained uncharacterized in annual killifishes. To confirm the quality of the genome assembly and assess the utility of the chromatin conformation data to interrogate 3D genome structure and gene regulation, we constructed a Hi-C contact map showing higher contact frequency within the 24 pearlfish chromosomes (Fig. 1d) than between chromosomes.
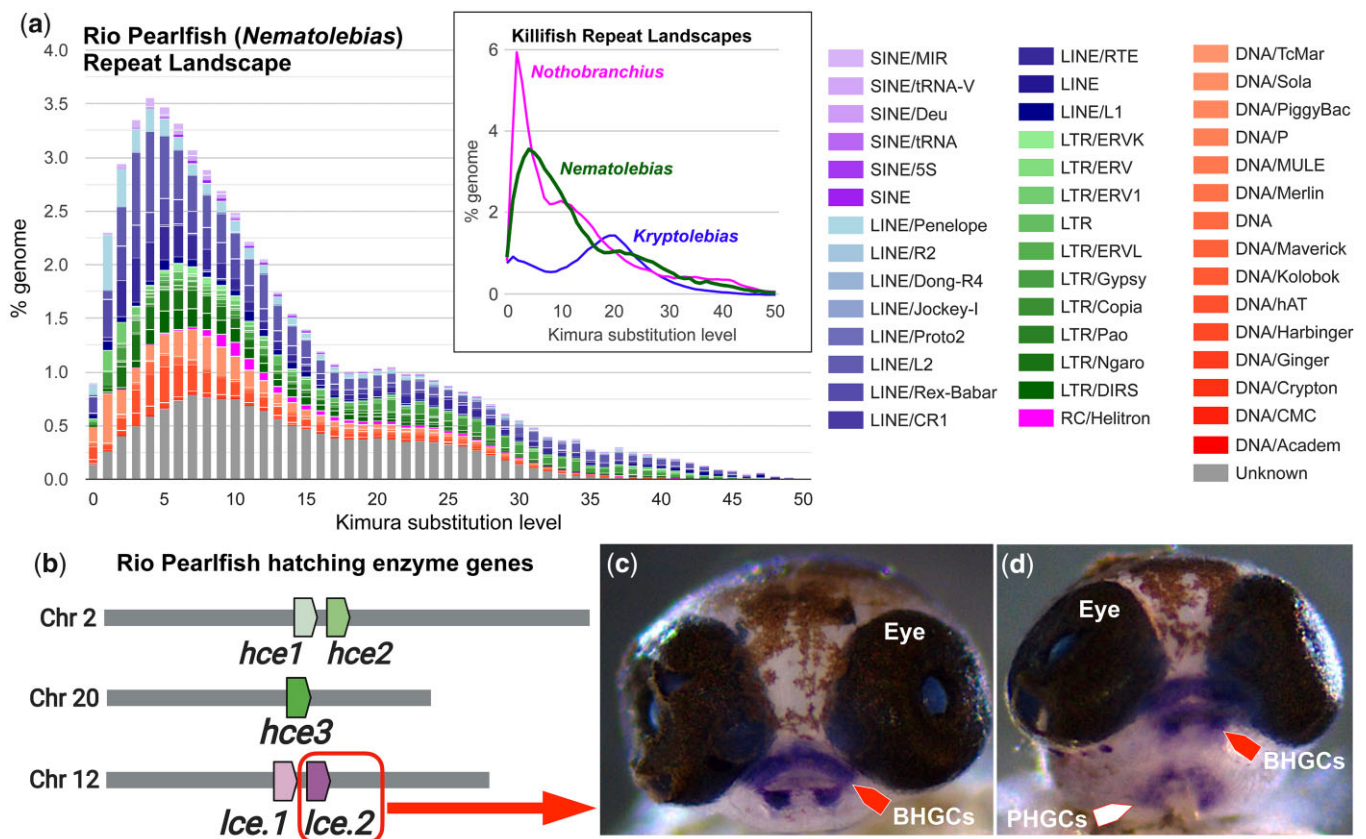
Fig. 2. Rio Pearlfish repeat landscape, hatching enzyme genes, and hatching gland location. a) Repeat landscape of mobile genetic elements in Rio Pearlfish showing a high repeat content with 2 peaks at Kimura distance 4 and 21. Insert: Total TE landscape among killifishes with independent, recent expansions in the convergently annual *Nothobranchius* (Cui *et al.* 2019) and *Nematolebias* (this study) compared to the nonannual *Kryptolebias* (Choi *et al.* 2020). b) Locations of 5 hatching enzyme genes in the Rio Pearlfish genome expressed during DIII. c, d) Wholemount RNA *in situ* hybridization for *lce.2* in DIII Rio Pearlfish embryos marking hatching gland cells (HGCs) identified in the buccal (BHGCs, red arrows) and pharyngeal (PHGCs, white arrow) cavities.

Using the genome sequence and gene annotations for Rio Pearlfish in synteny comparisons to another atherinomorph teleost, the medaka (separated by ∼85 million years), and the ostariophysian teleost zebrafish (separated by ∼224 million years), we reveal largely conserved synteny with these species (Fig. 2, e and f) across millions of years of teleost evolution (Hughes *et al.* 2018; Thompson *et al.* 2021a). We examined a TAD previously shown to be conserved from zebrafish to medaka (Nakamura *et al.* 2021) and found high frequency of contacts in the syntenic region between *rasa1a* and *mctp1a* in Rio Pearlfish liver tissue that resembles contact maps both in a medaka fibroblast cell line and zebrafish whole embryos (Fig. 1g). Hi-C analyses thus confirm the high-quality of our genome assembly as well as the strikingly conserved nature of 3D genome interactions across teleost evolution, developmental stages, and among cell and tissue types.

Repeat content and transposable element landscape

Transposable elements (TEs) are hypothesized to generate novel genetic substrate for adaptations (Casacuberta and González 2013; Feiner 2016; Esnault *et al.* 2019). Some annual killifish species have expanded TE content compared to nonannual relatives (Cui *et al.* 2019), and the link between TEs, aging, and human diseases (Bravo *et al.* 2020) coupled with the rapid senescence of annual killifishes highlights the importance of examining the Pearlfish "mobilome." We found that the Rio Pearlfish genome is highly repetitive with a repeat content of ∼57% (Fig. 2a and

Table 1; Supplementary Table 10) which is substantially elevated compared to a nonannual member of the same South American family, *K. marmoratus*, with a repeat content of around ∼27% (Rhee *et al.* 2017; Choi *et al.* 2020). Similarly, African annual *Nothobranchius* killifishes have higher TE repeat content than their nonannual relatives (Cui *et al.* 2019). This pattern might be the result of adaptation to extreme environments as animals, fungi, and plants have co-opted TEs for environmental adaptations to harsh conditions (Casacuberta and González 2013; Esnault *et al.* 2019) and TEs may play roles in vertebrate adaptive radiations (Feiner 2016). Our findings further highlight the expanded repeat content in annual killifish genomes and the Pearlfish genome provides novel resources to study the role of mobile DNA in extremophiles.

Hatching enzyme gene expression and hatching gland location

While hatching from the egg is a critical time point during animal development, little is known about its genetic regulation and the integration of environmental cues. Additionally, development of HGCs is dynamic among fishes (Inohaya *et al.* 1995; 1997) as they migrate and localize in different anatomical locations in different species (Korwin-Kossakowski 2012; Shimizu *et al.* 2014; Nagasawa *et al.* 2016). Pinpointing HGC location in seasonal killifishes is necessary for understanding the regulation of hatching in extreme environments.

Rio Pearlfish is a tractable model for studying hatching regulation since hatching is easily induced in this species by exposing DIII embryos to water (Thompson and Ortí 2016). Thus, we examined the hatching enzyme gene repertoire and HGC locations in Pearlfish. We identified five expressed hatching enzyme genes (Fig. 2b, 3 *hce* and 2 *lce* genes) upon mapping DIII mRNA reads from Thompson and Ortí (2016) to our reference genome assembly. We annotated *hce1* and *hce2* on chromosome 2 (corresponding to NCBI genes LOC119423801, LOC119423789), and *hce3* on chromosome 20 (LOC119426643) and the adjacent *lce.1* and *lce.2* genes (LOC119418488, LOC119418489) on chromosome 12 that are species-specific tandem duplicates (Fig. 2b) supported by transcript evidence (Thompson and Ortí 2016). Using whole mount RNA *in situ* hybridization for *lce.2* in DIII embryos, we identified HGC locations in the buccal and pharyngeal cavity in Rio Pearlfish (Fig. 2, c and d) similar to HGC localization in the related mummichog or Atlantic killifish (*Fundulus heteroclitus*) (Kawaguchi et al. 2005) and in medaka (Inohaya et al. 1995).

A pattern of expanded *hce* genes is also found in other Atherinomorph fishes like medaka. High choriolytic enzyme genes in this clade of teleosts have lost introns (Kawaguchi et al. 2010) and subfunctionalized postduplication with some *hce* genes performing better at higher or lower salinities in the euryhaline medaka *Oryzias javanicus* (Takehana et al. 2020) and the Atlantic killifish *F. heteroclitus* (Kawaguchi et al. 2013b). Furthermore, the duplication of the *lce* gene in Rio Pearlfish is an example of convergent evolution within teleosts with another *lce* duplication in stickleback fishes (Kawaguchi et al. 2013a). These findings underscore the commonality of hatching enzyme gene duplications in teleost fishes that provides a model system for studying convergent gene duplications and functional divergence by sub- and neofunctionalizations.

Conclusions

Our chromosome-level, dually annotated genome assembly of the Rio Pearlfish provides a valuable comparative genomics resource strengthening the utility of killifishes for studying aging, suspended animation, and response to environmental stress. The Rio Pearlfish is an emerging "Extremo" Eco-Evo-Devo research organism, and this reference genome will be a substrate for future functional genetic and multiomic approaches exploring how organisms integrate developmental and environmental cues to adapt to extreme environmental conditions in a changing world.

Data availability

The genome sequence, annotation, and sequence read data are available on NCBI under accession GCA_014905685.2 and Bioproject PRJNA560526. The genome assembly and annotation has also been integrated to the University of California Santa Cruz Genome Browser (https://hgdownload.soe.ucsc.edu/hubs/fish/index.html). The MAKER genome annotation is available on github (https://github.com/AndrewWT/NematolebiasGenomics).

Supplemental material is available at *G3* online.

Acknowledgments

Funding

Conflicts of interest

None declared.

Literature cited

Aluru N, Karchner SI, Franks DG, Nacci D, Champlin D, Hahn ME. Targeted mutagenesis of aryl hydrocarbon receptor 2a and 2b genes in Atlantic killifish (*Fundulus heteroclitus*). Aquat Toxicol. 2015;158:192–201.

Amemiya CT, Alföldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N, *et al.* The African coelacanth genome provides insights into tetrapod evolution. Nature. 2013;496(7445):311–316.

Bowman MJ, Pulman JA, Liu TL, Childs KL. A modified GC-specific MAKER gene annotation method reveals improved and novel gene predictions of high and low GC content in *Oryza sativa*. BMC Bioinformatics. 2017;18(1):15.

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, *et al.* The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 2016;48(4):427–437.

Bravo JI, Nozownik S, Danthi PS, Benayoun BA. Transposable elements, circular RNAs and mitochondrial transcription in age-related genomic regulation. Development. 2020;147(11):1–18.

Campbell, MS, Holt, C, Moore, B, Yandell, M. Genome annotation and curation using MAKER and MAKER-P. Curr Protoc Bioinform. 2014;48(1):4–11.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18(1):188–196.

Carter CA, Wourms JP. Naturally occurring diblastodermic eggs in the annual fish *Cynolebias*: implications for developmental regulation and determination. J Morphol. 1993;215(3):301–312.

Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Mol Ecol. 2013;22(6):1503–1517.

Choi BS, Park JC, Kim MS, Han J, Kim DH, Hagiwara A, Sakakura Y, Hwang UK, Lee BY, Lee JS. The reference genome of the selfing fish *Kryptolebias hermaphroditus*: identification of phases I and II detoxification genes. Comp Biochem Physiol Part D Genomics Proteomics. 2020;35:100684.

Costa WJEM, Amorim PF, Aranha GN. Species limits and DNA barcodes in *Nematolebias*, a genus of seasonal Killifishes threatened with extinction from the Atlantic Forest of South-Eastern Brazil,

with description of a New Species (Teleostei Rivulidae). Ichthyol Explor Freshwaters. 2014;24:225–236.

Costa WJEM. The neotropical seasonal fish genus *Nematolebias* (Cyprinodontiformes Rivulidae Cynolebiatinae) taxonomic revision with description of a new species. Ichthyol Explor Freshwaters. 2002;1:41–52.

Cui R, Medeiros T, Willemsen D, Iasi LNM, Collier GE, Graef M, Reichard M, Valenzano DR. Relaxed selection limits lifespan by increasing mutation load. Cell. 2019;178(2):385–399.

Deyts C, Candal E, Joly JS, Bourrat F. An automated in situ hybridization screen in the medaka to identify unknown neural genes. Dev Dyn. 2005;234(3):698–708.

Drost HG, Gabel A, Grosse I, Quint M. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. Mol Biol Evol. 2015;32(5):1221–1231.

Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1):14.

Esnault C, Lee M, Ham C, Levin HL. Transposable element insertions in fission yeast drive adaptation to environmental stress. Genome Res. 2019;29(1):85–95.

Feiner N. Accumulation of transposable elements in hox gene clusters during adaptive radiation of anolis lizards. Proc R Soc B Biol Sci. 2016;283(1840):20161555.

Furness AI, Reznick DN, Springer MS, Meredith RW. Convergent evolution of alternative developmental trajectories associated with diapause in African and South American killifish. Proc R Soc B. 2015;282(1802):20142189.

Hara Y, Yamaguchi K, Onimaru K, Kadota M, Koyanagi M, Keeley SD, Tatsumi K, Tanaka K, Motone F, Kageyama Y, *et al.* Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. Nat Ecol Evol. 2018;2(11):1761–1771.

Harel I, Benayoun BA, Machado B, Singh PP, Hu C-K, Pech MF, Valenzano DR, Zhang E, Sharp SC, Artandi SE, *et al.* A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. Cell. 2015;160(5):1013–1026.

Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491.

Hong CS, Saint-Jeannet JP. Xhe2 is a member of the astacin family of metalloproteases that promotes *Xenopus* hatching. Genesis. 2014;52(12):946–951.

Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, Arcila D, Betancur-R R, Li C, Becker L, *et al.* Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc Natl Acad Sci U S A. 2018;115(24):6249–6252.

Inohaya K, Yasumasu S, Araki K, Naruse K, Yamazaki K, Yasumasu I, Iuchi I, Yamagami K. Species-dependent migration of fish hatching gland cells that commonly express astacin-like proteases in common. Dev Growth Differ. 1997;39(2):191–197.

Inohaya K, Yasumasu S, Ishimaru M, Ohyama A, Iuchi I, Yamagami K. Temproral and spatial patterns of gene expression for the hatching enzyme in the teleost embryo. Dev Biol. 1995;171(2):374–385.

Jurka J. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 2000;16(9):418–420.

Kawaguchi M, Hiroi J, Miya M, Nishida M, Iuchi I, Yasumasu S. Intron-loss evolution of hatching enzyme genes in teleostei. BMC Evol Biol. 2010;10:1–10.

Kawaguchi M, Takahashi H, Takehana Y, Naruse K, Nishida M, Yasumasu S. Sub-functionalization of duplicated genes in the

evolution of nine-spined stickleback hatching enzyme. J Exp Zool B Mol Dev Evol. 2013a;320(3):140–150.

Kawaguchi M, Yasumasu S, Hiroi J, Naruse K, Inoue M, Iuchi I. Evolution of teleostean hatching enzyme genes and their paralogous genes. Dev Genes Evol. 2006;216(12):769–784.

Kawaguchi M, Yasumasu S, Shimizu A, Hiroi J, Yoshizaki N, Nagata K, Tanokura M, Iuchi I. Purification and gene cloning of *Fundulus heteroclitus* hatching enzyme: a hatching enzyme system composed of high choriolytic enzyme and low choriolytic enzyme is conserved between two different teleosts, *Fundulus heteroclitus* and medaka *Oryzias latipes*. FEBS J. 2005;272(17):4315–4326.

Kawaguchi M, Yasumasu S, Shimizu A, Kudo N, Sano K, Iuchi I, Nishida M. Adaptive evolution of fish hatching enzyme: one amino acid substitution results in differential salt dependency of the enzyme. J Exp Biol. 2013b;216(Pt 9):1609–1615.

Kelley JL, Yee MC, Brown AP, Richardson RR, Tatarenkov A. The genome of the self-fertilizing mangrove rivulus fish, *Kryptolebias marmoratus*: a model for studying phenotypic plasticity and adaptations to extreme environments. Genome Biol Evol. 2016;8(7):2145–2154.

Korwin-Kossakowski M. Fish hatching strategies: a review. Rev Fish Biol Fisheries. 2012;22(1):225–240.

Krefting J, Andrade-Navarro MA, Ibn-Salem J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. BMC Biol. 2018;16(1):1–12.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–1760.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289–293.

Lyons E, Freeling M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. 2008;53(4):661–673.

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 2021;49(D1):D394–D403.

Myers GS. Annual fishes. Aquarium J. 1952;23:125–141.

Myers GS. Studies on South American freshwater fishes I. Stanford Ichthyol Bull. 1942;2:89–114.

Nagasawa T, Kawaguchi M, Yano T, Sano K, Okabe M, Yasumasu S. Evolutionary changes in the developmental origin of hatching gland cells in basal ray-finned fishes. Zool Sci. 2016;33(3):272–281.

Nakamura R, Motai Y, Kumagai M, Wike CL, Nishiyama H, Nakatani Y, Durand NC, Kondo K, Kondo T, Tsukahara T, *et al.* CTCF looping is established during gastrulation in medaka embryos. Genome Res. 2021;31(6):968–980.

Nishimura O, Hara Y, Kuraku S. GVolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics. 2017;33(22):3635–3637.

Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23(9):1061–1067.

Podrabsky JE, Hand SC. The bioenergetics of embryonic diapause in an annual Killifish, *Austrofundulus limneaus*. J Exp Biol. 1999;202(19):2567–2580.

Podrabsky JE, Lopez JP, Fan TWMM, Higashi R, Somero GN. Extreme anoxia tolerance in embryos of the annual killifish *Austrofundulus*

limnaeus: insights from a metabolomics analysis. J Exp Biol. 2007; 210(Pt 13):2253–2266.

Podrabsky JE, Menze MA, Hand SC. Long-term survival of anoxia despite rapid ATP decline in embryos of the annual killifish *Austrofundulus limnaeus*. J Exp Zool A Ecol Genet Physiol. 2012; 317(8):524–532.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 2016;26(3):342–350.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun. 2018;9(1):189.

Reichwald K, Petzold A, Koch P, Downie BR, Hartmann N, Pietsch S, Baumgart M, Chalopin D, Felder M, Bens M, *et al.* Insights into sex chromosome evolution and aging from the genome of a short-lived fish. Cell. 2015;163(6):1527–1538.

Rhee JS, Choi BS, Kim J, Kim BM, Lee YM, Kim IC, Kanamori A, Choi IY, Shartl M, Lee JS. Diversity, distribution, and significance of transposable elements in the genome of the only selfing hermaphroditic vertebrate *Kryptolebias marmoratus*. Sci Rep. 2017;7(1):1–10.

Ruijter JM, Creuwels LAJM. The ultrastructure of prolactin cells in the annual cyprinodont *Cynolebias whitei* during its life cycle. Cell Tissue Res. 1988;253(2):477–483.

Ruijter JM, Van Kemenade JAM, Bonga SEW. Environmental influences on prolactin cell development in the cyprinodont fish, *Cynolebias whitei*. Cell Tissue Res. 1984;238(3):595–600.

Ruijter JM. Development and aging of the teleost pituitary: qualitative and quantitative observations in the annual cyprinodont *Cynolebias whitei*. Anat Embryol (Berl). 1987;175(3):379–386.

Sano K, Kawaguchi M, Watanabe S, Yasumasu S. Neofunctionalization of a duplicate hatching enzyme gene during the evolution of teleost fishes. BMC Evol Biol. 2014;14:221.

Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff J-N, Lesch K-P, *et al.* The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. Nat Genet. 2013;45(5):567–572.

Schoots AFM, Ruijter JM, van Kemenade JAM, Denucé JM. Immunoreactive prolactin in the pituitary gland of cyprinodont fish at the time of hatching. Cell Tissue Res. 1983;233(3):611–618.

Shimizu D, Kawaguchi M, Sano K, Yoshizaki N, Shimizu D, Fujinami Y, Noda T, Yasumasu S. Comparison of hatching mode in pelagic and demersal eggs of two closely related species in the order Pleuronectiformes. Zoolog Sci. 2014;31(11):709–715.

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.

Simpson BRC. The phenology of annual killifish. Symp Zool Soc Lon. 1979;44:243–261.

Smit AFA, Hubley R. RepeatModeler Open-1.0; 2008. http://www.repeatmasker.org.

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

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9): 1312–1313.

Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–2690.

Takehana Y, Zahm M, Cabau C, Klopp C, Roques C, Bouchez O, Donnadieu C, Barrachina C, Journot L, Kawaguchi M, *et al.*

Genome sequence of the euryhaline Javafish Medaka, *Oryzias javanicus*: a small aquarium fish model for studies on adaptation to salinity. G3 (Bethesda). 2020;10(3):907–915.

Thibaud-Nissen F, DiCuccio M, Hlavina W, Kimchi A, Kitts PA, Murphy TD, Pruitt KD, Souvorov A. P8008 the NCBI eukaryotic genome annotation pipeline. J Anim Sci. 2016;94(Suppl 4):184.

Thompson AW, Black AC, Huang Y, Shi Q, Furness AI, *et al.* Deterministic shifts in molecular evolution correlate with convergence to annualism in killifishes. BioRxiv, 2021a, https://doi.org/10.1101/2021.08.09.455723

Thompson AW, Furness AI, Stone C, Rade C, Ortí G. Microanatomical diversification of the zona pellucida in aplochelioid killifishes. J Fish Biol. 2017a;91(1):126–143.

Thompson AW, Hawkins MB, Parey E, Wcisel DJ, Ota T, Kawasaki K, Funk E, Losilla M, Fitch OE, Pan Q, *et al.* The bowfin genome illuminates the developmental evolution of ray-finned fishes. Nat Genet. 2021b;53(9):1373–1384.

Thompson AW, Hayes A, Podrabsky JE, Ortí G. Gene expression during delayed hatching in fish-out-of-water. Ecol. Genet. Genomics. 2017b;3–5:52–59.

Thompson AW, Ortí G. Annual killifish transcriptomics and candidate genes for metazoan diapause. Mol Biol Evol. 2016;33(9): 2391–2395.

Valenzano DR, Benayoun BA, Singh PP, Zhang E, Etter PD, Hu C-K, Clément-Ziza M, Willemsen D, Cui R, Harel I, *et al.* The African Turquoise Killifish genome provides insights into evolution and genetic architecture of lifespan. Cell. 2015;163(6):1539–1554.

Valenzano DR, Sharp SC, Brunet A. Transposon-mediated transgenesis in the short-lived African Killifish *Nothobranchius furzeri*, a vertebrate model for aging. G3 (Bethesda). 2011;1(7):531–538.

Van Post A. Vergleichende Untersuchungen der Chromosomenzahlen bei Süßwasser Teleosteern. Zeitschrift Für Zool Syst Evol. 1965;3(1–2):47–93.

Vrtílek M, Žák J, Pšenička M, Reichard M. Extremely rapid maturation of a wild African annual fish. Curr Biol. 2018;28(15):R822–R824.

Wagner JT, Singh PP, Romney AL, Riggs CL, Minx P, Woll SC, Roush J, Warren WC, Brunet A, Podrabsky JE, *et al.* The genome of *Austrofundulus limnaeus* offers insights into extreme vertebrate stress tolerance and embryonic development. BMC Genomics. 2018;19(1):21.

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Corrigendum: direct determination of diploid genome sequences. Genome Res. 2018;28(4):606.1–767.

Wourms JP. Annual fishes. In: Methods in Developmental Biology. New York: Thomas and Crowell Company; 1967. p. 123–137.

Wourms JP. The developmental biology of annual fishes I. Stages in the normal development of *Austrofundulus myersi* Dahl. J Exp Zool. 1972a;182(2):143–168.

Wourms JP. The developmental biology of annual fishes II. Naturally occurring dispersion and reaggregation of blastomeres during the development of annual fish eggs. J Exp Zool. 1972b;182(2):169–200.

Wourms JP. The developmental biology of annual fishes III. Preembryonic and embryonic diapause of variable duration in the eggs of annual fishes. J Exp Zool. 1972c;182(3):389–414.

Yamagami K. Mechanisms of hatching in fish. In: Fish Physiology. Cambridge, Massachusetts: Academic Press; 1988. p. 447–499.

Yasumasu S, Yamada K, Akasaka K, Mitsunaga K, Iuchi I, Shimada H, Yamagami K. Isolation of cDNAs for LCE and HCE, two constituent proteases of the hatching enzyme of *Oryzias latipes*, and concurrent expression of their mRNAs during development. Dev Biol. 1992;153(2):250–258.

Communicating editor: A. Whitehead