scientific reports

Check for updates

OPEN

Description and phylogenetic implications of a de novo mitochondrial genome of *Rhinichthys atratulus* (Teleostei: Leuciscidae) from Connecticut

Timothy S. Earley[1,4]✉, Naomi Whitlock[2,3,4], Samuel J. Taylor[2], Antonio Machado-Allison[1] & Barry Chernoff[1,2,3]

We present a de novo mitogenome assembly from a specimen of *Rhinichthys atratulus*, the Eastern Blacknose Dace, collected in the Connecticut River drainage. *R. atratulus* is a fish species widely distributed across Atlantic slope drainages from Nova Scotia, Canada to the Roanoke River Drainage, Virginia, United States. The assembly has a total length of 16,646 bp; consists of 13 protein-coding genes, 22 tRNA genes, 2 rRNA genes, and a 974 bp D-loop; and has a GC content of 45.7%. We performed two phylogenetic analyses using the de novo assembly and 28 additional cyprinoid mitochondrial genomes: with and without the D-loop. The resulting phylogenetic trees showed the same branching patterns. However, inclusion of the D-loop resulted in nodes with equal or greater support. The trees largely match previously published analyses, while offering new insights into relationships among leuciscid genera. Our findings indicate that the inclusion of the D-loop in mitogenome phylogenetic analyses can help improve bootstrap support at otherwise unresolved nodes.

Mitochondrial genomes are a useful tool for understanding evolution at different levels of biological organization, including ecological effects on populations[1], divergence patterns among closely related species[2], and elucidating phylogenetic relationships[3]. The reason for the versatility of the mitogenome is that the mitochondrial genes and regulatory regions such as the control region, also known as D-loop—herein used in the broad sense to refer to the entire non-coding region between the tRNA-Pro and tRNA-Phe genes[4]—have been shown to evolve at different rates[5–8]. Thus, the relative ease and affordability of generating mitogenomes makes them an appealing tool for comparative population genomics and phylogenomics.

Here we describe the structure and content of a new high-quality mitochondrial genome of the Eastern Blacknose Dace, *Rhinichthys atratulus.* Mitogenomic data have revealed important insights into phylogenetic relationships of fishes[3,9]. An important consideration is whether to include D-loop sequence as part of the data set. In some prior studies, D-loop sequences have successfully contributed important information about relationships[10,11]. Other studies, however, have excluded D-loop sequence from analyses because of length variation or alignment problems[1,12,13]. In this paper, we illustrate the potential value of including the D-loop sequence in mitogenomic datasets, even when some sequence length variation exists across sampled taxa.

Osborne et al.[12] used mitogenomes without the D-loop to estimate phylogenetic relationships among a small set of cyprinoid fishes (minnows), rooting their tree with two outgroup species from the family Catostomidae. We expanded the number of ingroup cyprinoid terminal taxa to 27: two cyprinids and 25 leuciscids. We included our de novo mitogenome of *R. atratulus* in the study group in order to determine if *R. atratulus* is sister to *R. obtusus* as demonstrated by Kraczkowski and Chernoff[14] (see below). We compared our phylogenetic results to those of Osborne et al.[12] and illustrated the potential importance of including sequence information from the D-loop. Furthermore, we compared our results to the comprehensive study of cyprinoid species with 358 leucisid taxa based upon two mitochondrial (*cytb*, *COI*) and two nuclear (*rag1, IRBP*) genes[15].

[1]Bailey College of the Environment, Wesleyan University, Middletown, CT, USA. [2]Department of Biology, Wesleyan University, Middletown, CT, USA. [3]Department of Earth & Environmental Sciences, Wesleyan University, Middletown, CT, USA. [4]Timothy S. Earley and Naomi Whitlock contributed equally to this work. ✉email: tearley@wesleyan.edu

Study material and species identification

Eastern Blacknose Dace, *Rhinichthys atratulus* Hermann (1804), was obtained from a feeder stream to the Coginchaug River, a small tributary of the Connecticut River in central Connecticut, USA. The species generally inhabits medium to small streams and rivers, preferring riffles and rocky habitats over sand or clay[14]. Kraczkowski and Chernoff[14] demonstrated that *R. atratulus* is found in Atlantic slope drainages from Nova Scotia to the Roanoke River drainage in Virginia, and in southeastern tributaries to Lake Ontario.

The putative *R. atratulus* tissues for Schönhuth et al.[15,16] and the mitogenome NC_033943.1 are from specimens collected together in Russell Creek, northeast of Tazewell, Tennessee, a tributary of the Powell River, Tennessee, in the Clinch River and Tennessee River drainages. The specimens are catalogued in the University of Alabama Ichthyology Collection (UAIC 9850.01). We examined the specimens in UAIC 9850.01, and determined that they are Western Blacknose Dace, *Rhinichthys obtusus* Agassiz (1854). *R. atratulus* is not known from the Tennessee River basin[14]. It should be noted that the first paper by Schönhuth et al.[16] (2012) and the deposition of the sequence in GenBank (2011) predate the removal of *R. obtusus* from the synonymy of *R. atratulus* by Kraczkowski and Chernoff[14] (2014).

Results

The *R. atratulus* mitogenome assembly has a length of 16,646 bases and is composed of 13 protein-coding genes, 22 tRNA genes, 2 rRNA genes, and a 974 bp D-loop (Fig. 1), with a GC content of 45.7%. The assembly is supported by a mean read depth of 1677.481 (Fig. S1). Overall, the nucleic acids are relatively consistent across the sampled taxa (Fig. 2).

Among all mitogenomes examined, GC content comprises less than 50% (except for *Meda fulgida*) because guanine comprises less than 15% of the mitogenome. Thymines are greater than 25% of all the nucleic acids and more common than adenines, which are almost uniformly 25% (Fig. 2).

The D-loop of our *R. atratulus* assembly represented the median length among the taxa at 974 bp and was close to the mean length of 977.448 bp (Table 1). D-loop lengths varied considerably among sampled taxa, ranging from 890 bp (*Oregonichthys crameri*) to 1135 bp (*Tiaroga cobitis*). This large variation in D-loop length (e.g., the 245 bp difference between the *Oregonichthys* and *Tiaroga*) could potentially bias phylogenetic trees. Our results did not reflect this; the D-loop contained meaningful signals and increased bootstrap support at a number of nodes.

There was 100% consensus of the branching patterns for the phylogenetic trees that we generated, both with and without the D-loop sequences (Fig. 3). Both phylogenetic estimates demonstrated that ingroup cyprinoids and leuciscids were monophyletic, and that *R. atratulus* is most closely related to a monophyletic *R. obtusus*. Overall, the tree with the D-loop exhibited better bootstrap support than the tree that did not include D-loop sequences. Of the 26 nodes present in both trees, four nodes in the tree lacking D-loop sequence were poorly supported (i.e., bootstrap support < 70%). Inclusion of the D-loop reduced the number of poorly supported nodes to two. The support for the monophyly of *Semotilus, Meda,* and *Tribolodon* jumped from 50.2% to 77.7% (Fig. 3), and support for the monophyly of *Cyprinella* and two species of *Hybognathus* increased from 22 to 72.2% (Fig. 3).

Discussion and conclusions

Mitochondrial genomes are widely used to investigate biological questions at different levels in the hierarchy of life—from population-level questions[17] to species-level relationships[3] and to higher-order phylogenetic analyses[18]. Recent advances in genomic sequencing have reduced the time and cost of obtaining mitogenomes. In this paper we introduce a new mitogenome, *Rhinichthys atratulus*, to the growing list of mitogenomes available. Our new mitogenome is important because it comes from the exact locality of the neotypic series for the species—thus, if there is an *R. atratulus*, it comes from the Coginchaug River at Wadsworth Falls State Park.

We generated two phylogenetic trees from a sample of 29 cyprinoid fishes with and without sequences from the D-loop. Although the D-loop sequences required careful alignment, these sequences provided important support for nodes on the tree; the number of poorly supported nodes were halved (from 4 to 2). Osborne et al.[12] omitted the D-loop from their analyses due to concerns over variation in D-loop length across sampled taxa; however, our results suggest that this variation can carry phylogenetically useful information without reducing support elsewhere. Indeed, the species with the shortest and longest D-loop regions—*O. crameri* and *T. cobitis*, respectively—are sister taxa with very strong node support in both of our trees (99.8%, Fig. 3) despite the 245 bp difference between their respective D-loops.

We compared our phylogenetic results with the findings by Schönhuth et al.[15], Schönhuth et al.[16], and Osborne et al.[12]. Schönhuth et al.[15] produced three phylogenetic trees from: (i) two mitochondrial genes; and (ii) two nuclear genes; and (iii) combined mitochondrial and nuclear genes. Schönhuth et al.[16] produced a similar set of trees, but instead used only one mitochondrial gene and three nuclear genes. Our study and those of Osborne et al.[12] and Schönhuth et al.[16] included many fewer taxa than did Schönhuth et al.[15] (27, 21, 71, and 410, respectively). Our results were largely congruent with these studies. An important finding of our study is that *R. atratulus* is the sister group to *R. obtusus* (Fig. 3); the former species was not included in the other studies (see *Study Material and Species Identification* above for further details).

There are three areas of our tree that differ from Osborne et al.[12], Schönhuth et al.[15], or Schönhuth et al.[16]. The first concerns the species relationships within a clade containing *Hybognathus amarus*, *H. nuchalis*, *Cyprinella lutrensis*, and *Pimephales promelas*. Our results had *C. lutrensis* as sister to the *Hybognathus* species clade with *P. promelas* the next closest taxon. Osborne et al.[12] alternatively estimated that *P. promelas* was the sister group to *Hybognathus* with *C. lutrensis* their nearest relative. Lastly, Schönhuth et al.[15] had *C. lutrensis* and *P. promelas* as

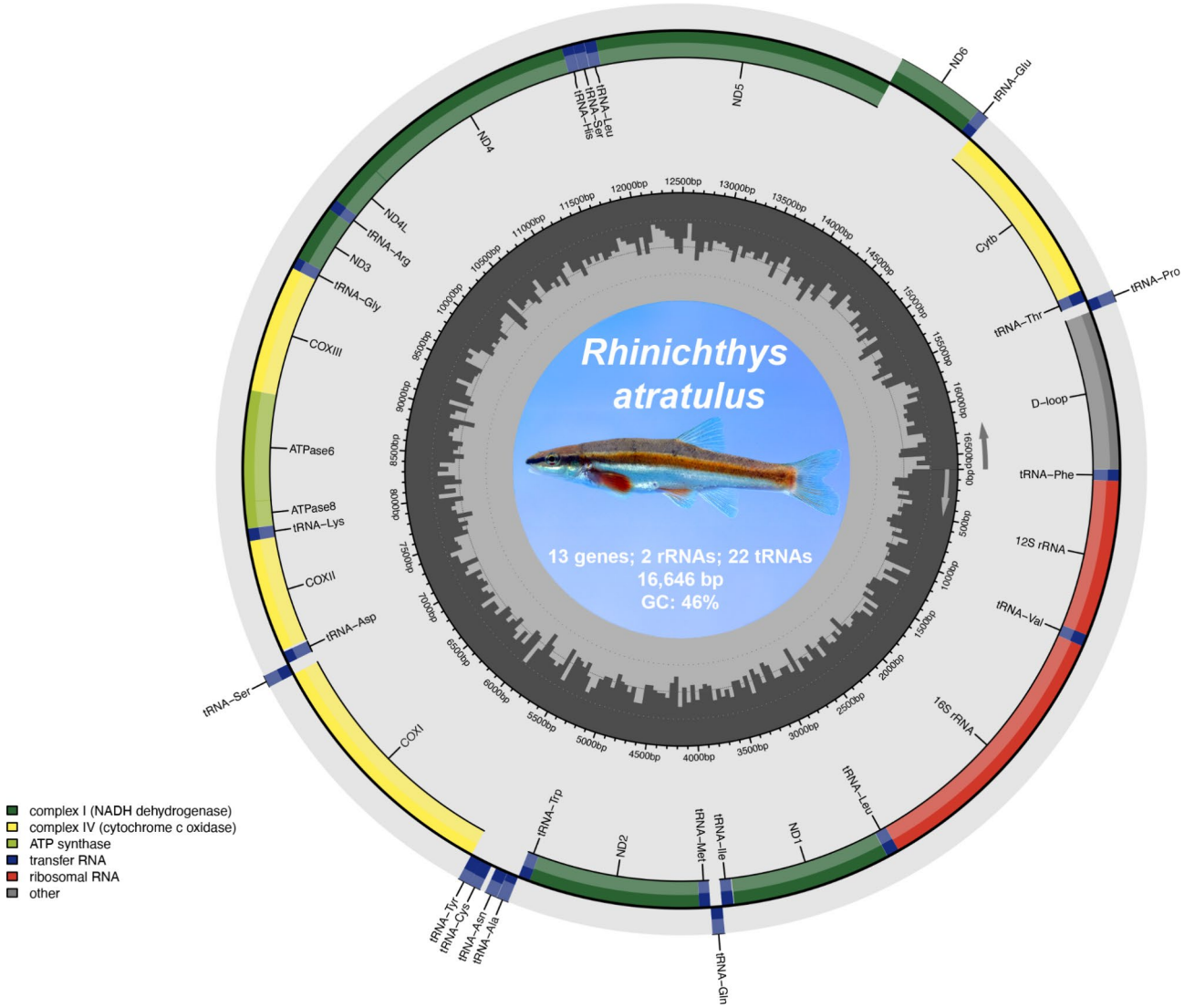

Fig. 1. Annotation of our de novo *R. atratulus* mitogenome assembly using MitoFish, visualized with Chloroplot (Zheng et al., 2020). The innermost ring represents the GC content of a given 100 bp interval of the sequence, with the subdivided rings representing increasing 25% intervals. The arrows in this section correspond to the direction of transcription for the genes and RNAs displayed in the outer ring, with the inner features transcribed clockwise (forward strand) and the outer features transcribed counterclockwise (reverse strand). Features are color coded by their product, per the key. The inset photograph is of a male *R. atratulus* individual in breeding colors. This individual was collected and photographed by the authors from the same locality as the sequenced individual in May 2023.

sister taxa with *Hybognathus* as sister to this clade. There is only medium support for each of these hypotheses in their respective studies.

A second difference is that our mitogenome trees have *Tiaroga* and *Oregonichthys* as most closely related to *Rhinichthys* (Fig. 3), in agreement with Schönhuth et al.[15]. However, the trees of Schönhuth et al.[16] with a different set of genes and taxa than Schönhuth et al.[15] suggested that *Tiaroga* is related to *Rhinichthys* at a more inclusive level. Our findings lend further support to the relationships described in Schönhuth et al.[15].

The third difference involves the relationship of *Erimystax* to *Exoglossum* and other pogonichthyines. *Exoglossum* is positioned as the sister to the non-*Mylocheilus*-clade[16] pogonichthyines in Schönhuth et al.[15,16] and our study. In our study *Erimystax* is sister to *Exoglossum* with 100% bootstrap support and with an exceedingly short branch length joining them (Fig. 3). In Schönhuth et al.[15] these genera are distantly related, and *Erimystax* is the sister group to *Phenacobius*; a result that occurs in their mitochondrial, nuclear, and combined trees.

The differences noted above can be due to several factors. The most obvious is taxon sampling, especially in relation to the potential problems of long branch attraction[19,20]. However, the branches in our analyses are not particularly long—especially in the case of *Erimystax* and *Exoglossum*, which have exceptionally short branches—and are well distributed. While we recognize that the superb sampling of Schönhuth et al.[15] is optimal and has the best chance of escaping this issue, the differences among our studies[12,15,16] may also have to do with

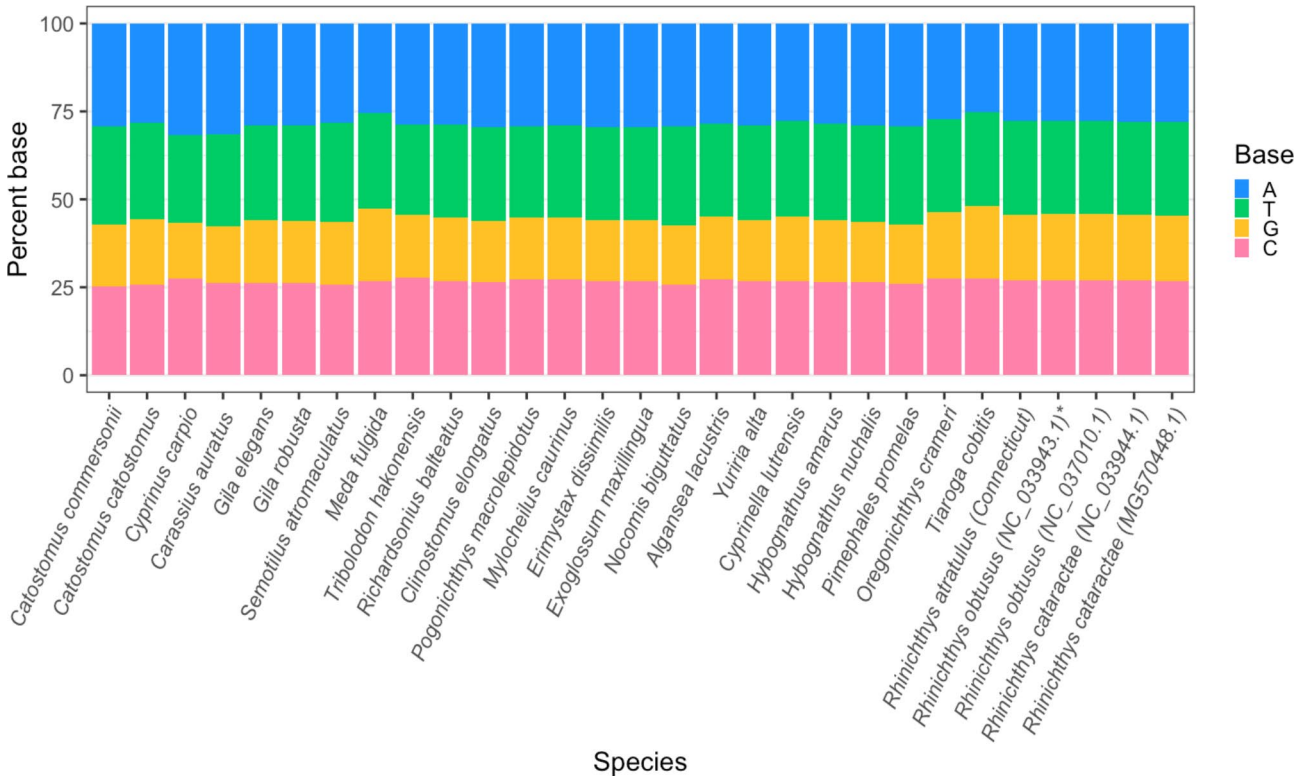

Fig. 2. Comparison of nucleic acid composition (as percent of total sequence, y-axis) between each mitochondrial genome (x-axis). The differently colored sections of each bar represent the proportion of A (blue), T (green), G (gold), and C (pink) in each mitogenome. The species are ordered according to the phylogenetic results shown in Fig. 3. *This specimen (NC_033943.1) is labeled as *Rhinichthys atratulus* in the Genbank but is properly identified as *R. obtusus*[14]; see *Study Material and Species Identification* for further information.

phylogenetic signal. The complete mitogenome, including the D-loop, of the cypriniform fishes of our study is at least five times larger than the total sequence used in Schönhuth et al.[15]. Thus, the phylogenetic signal of the mitogenomes in our study may be very strong. Additional study to compare phylogenetic signal from nuclear genomes with mitogenomes will resolve this issue.

This paper demonstrated the importance of mitogenomic data in addressing phylogenetic relationships, and especially the inclusion of the D-loop sequence. Many other studies have benefitted from inclusion of D-loop sequences ranging from phylogenetic relationships to population level studies[10,11]. *R. atratulus* and *R. obtusus* have been demonstrated to be monophyletic, confirming previous work[14]. Our study has raised questions about the phylogenetic relationships of *Cyprinella, Hybognathus* and *Pimephales*; *Tiaroga* and *Rhinichthys*; and *Exoglossum* and *Erimystax*. The question of their relationships within the Pogonichthyinae will require additional taxa and potentially analyses of mitogenomes and nuclear genomes to resolve the issues.

Our introduction of a new *Rhinichthys atratulus* genome brings attention to ensuring that we collect mitogenomes from species at or near type localities in order to ensure correct taxonomic nomenclature. Furthermore, the mitogenome of *R. atratulus* is an important resource for our ongoing phylogeographic, population evolution, and conservation studies of the *R. atratulus—R. obtusus* species complex.

Materials and methods

The sequenced specimen was captured in a small tributary of the Coginchaug River at Wadsworth Falls State Park, Middletown, Middlesex County, Connecticut River Drainage, CT, USA, longitude-72.687042, latitude 41.535731. The fish was collected via electrofishing and euthanized in a solution of tricaine methanesulfonate (MS-222) at a concentration of 500 mg/L. Fish collection and handling was humane, ethical, and approved by state and university animal care and use committees (IACUC2017-1212-Chernoff-A); all methods were performed in accordance with the relevant guidelines and regulations, and this study conforms to ARRIVE reporting guidelines. We sent a single whole fish to Cantata Genome Services (Cantata Bio, LLC, Scotts Valley, CA, USA). Whole genomic DNA was extracted from the muscle tissue and sequenced via PacBio circular consensus sequencing (Pacific Biosciences of California Inc., Menlo Park, CA, USA), producing a total read length 169.1 Gbp. A chromosome-level genome was assembled de novo with HiFiasm[21] (v. 0.15.4-r347). Contamination was removed with BlobTools[22] (v. 1.1.1) using the NCBI nt database[23] as reference, while haplotigs and contig overlaps were removed with purge_dups[24] (v. 1.2.5). The mitogenome was isolated from the de novo assembly using Minimap2[25] (v. 2.26) by aligning against the *Rhinichthys cataractae* mitogenome (GenBank accession

Taxa	NCBI accession	Length (bp)	D-loop length (bp)
Algansea lacustris	NC_031565.1	16,719	1055
Carassius auratus	NC_006580.1	16,580	922
Catostomus catostomus	NC_037013.1	16,622	920
Catostomus commersonii	NC_008647.1	16,625	950
Clinostomus elongatus	NC_031572.1	16,650	975
Cyprinella lutrensis	AB070206.1	16,706	1049
Cyprinus carpio	NC_001606.1	16,575	927
Erimystax dissimilis	NC_031571.1	16,653	985
Exoglossum maxillingua	NC_037015.1	16,648	957
Gila elegans	MT364326.1	16,593	915
Gila robusta	NC_008105.1	16,595	915
Hybognathus amarus	MT364325.1	16,713	1049
Hybognathus nuchalis	NC_031567.1	16,712	1049
Meda fulgida	NC_028291.1	16,591	920
Mylocheilus caurinus	NC_013763.1	16,657	981
Nocomis biguttatus	NC_033924.1	16,650	982
Oregonichthys crameri	NC_066939.1	16,653	890
Pimephales promelas	NC_028087.1	16,709	1049
Pogonichthys macrolepidotus	NC_033942.1	16,649	977
Rhinichthys atratulus (Connecticut)	PP958238	16,646	974
Rhinichthys cataractae	MG570448.1	16,655	958
Rhinichthys cataractae	NC_033944.1	16,655	982
*Rhinichthys obtusus**	NC_033943.1	16,651	976
Rhinichthys obtusus	NC_037010.1	16,650	954
Richardsonius balteatus	NC_033945.1	16,646	972
Semotilus atromaculatus	NC_033946.1	16,623	947
Tiaroga cobitis	MT364327.1	16,802	1135
Tribolodon hakonensis	NC_018820.1	16,602	930
Yuriria alta	NC_031566.1	16,715	1051

Table 1. Mitogenome and D-loop lengths for selected taxa with corresponding accession numbers. The *Length* column lists the total length of the RefSeq in base pairs (bp), while the *D-loop length* column lists the length of the non-coding control region between the tRNA-Pro and tRNA-Phe genes (Nicholls and Minczuk, 2014). *This specimen (NC_033943.1) is labeled as *Rhinichthys atratulus* in the Genbank but is properly identified as *R. obtusus*[14]; see *Study Material and Species Identification* for further information.

MG570448.1) and extracting the aligned scaffold using seqtk[26] (v. 1.3). We validated the identity of the extracted mitogenome by running a BLASTn search[27] (v. 2.14.0) against the NCBI mitochondrial RefSeq database[28]. Read coverage depth was determined using Winnowmap2[29] (v. 2.03). Annotations were completed using MitoFish[30].

Phylogenetic analysis involved alignment of the de novo *R. atratulus* mitogenome with 28 additional complete mitogenomes sourced from the NCBI RefSeq database (see Table 1). We aligned complete mitochondrial genomes with Clustal Omega[31] (v. 1.2.3) using default parameters. Due to variation in D-loop length among the selected genomes, we ran an additional alignment with the D-loops removed. Phylogenetic trees were constructed in R[32] (v. 4.3.2) using tidyverse[33] (v. 2.0.0), ape[34] (v. 5.7–1), phangorn[35] (v. 2.11.1; Schliep, 2011), and data.table[36] (v. 1.15.0) packages. Using the phangorn function *modelTest*, we selected GTR + G(4) + I as the evolutionary model based on AIC values. Each phylogeny was constructed with 1000 bootstrap iterations. Trees were visualized using the R package ggtree[37] (v. 3.10.1). Sequence base composition was calculated using seqtk.

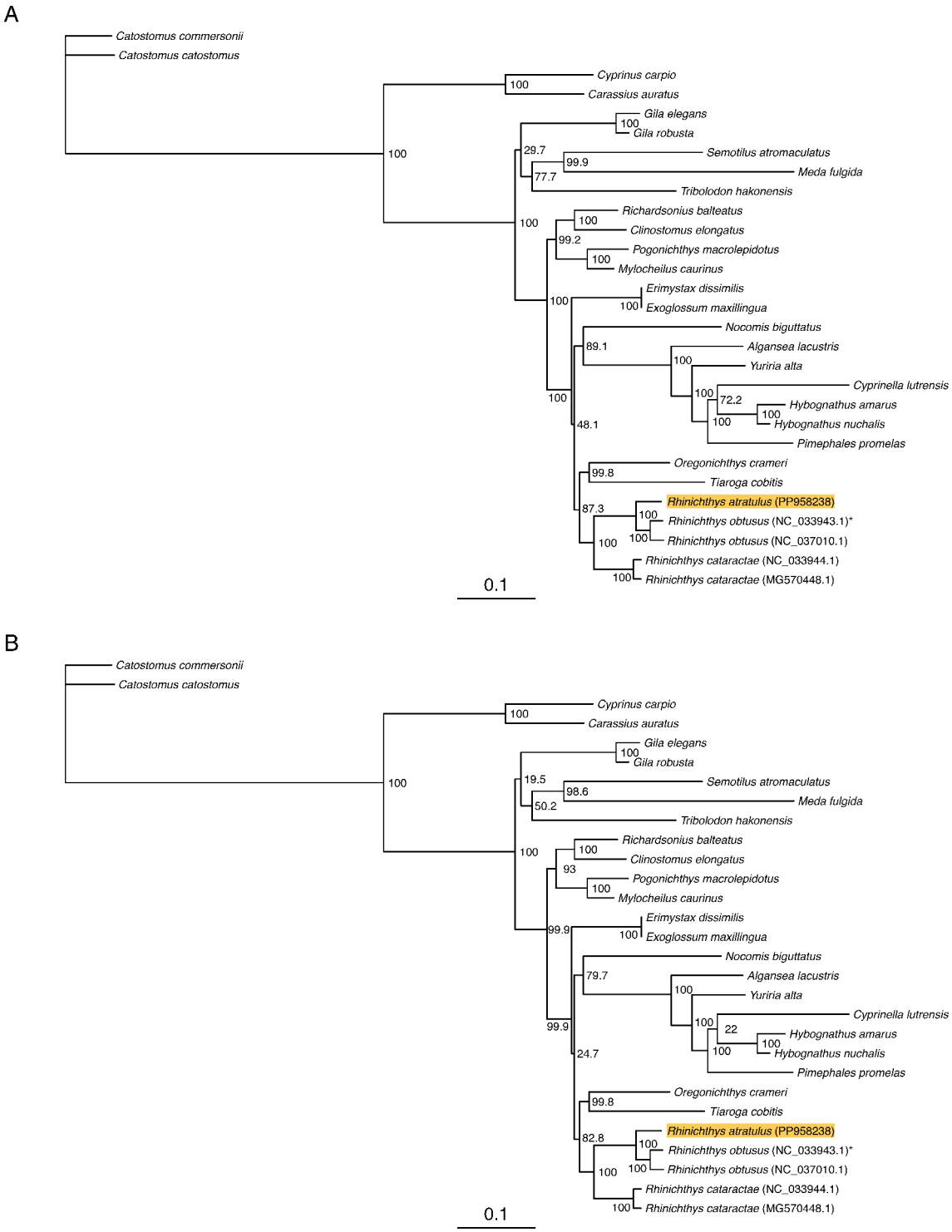

Fig. 3. Phylogenetic estimates from mitogenomes of 27 ingroup cyprinoid taxa (see Table 1 for complete list of NCBI accessions). The de novo *Rhinichthys atratulus* (PP958238) mitogenome is highlighted in gold. The trees were rooted at the two outgroups of the genus *Catostomus*. Trees were constructed (A) with D-loops and (B) without D-loops. Bootstrap values are displayed adjacent to each node. *This specimen (NC_033943.1) is labeled as *Rhinichthys atratulus* in the Genbank but is properly identified as *R. obtusus*[14]; see *Study Material and Species Identification* for further information.

Data availability

The de novo R. atratulus mitogenome assembly and annotations presented here are available in NCBI GenBank at accession PP958238. Sample reads and data are available through NCBI SRA SUB14590619, BioProject PRJNA1134187, and BioSample SAMN42395960. Supporting data for analyses are available at https://doi.org/10.25

438/wes02.c.7312162 and can be directly supplied upon written request to the authors. The sequenced specimen is in the Wesleyan University Ichthyology Collection as accession 000101; contact Barry Chernoff (bchernoff@wesleyan.edu) with any inquiries.

Received: 20 August 2024; Accepted: 11 December 2024
Published online: 30 December 2024

References

1. Osborne, M. J., Barela Hudgell, M. A., Caeiro-Dias, G. & Turner, T. F. The complete mitochondrial genomes of two imperiled species endemic to the Southwestern United States: Peppered Chub (*Macrhybopsis tetranema*) and Gila Trout (*Oncorhynchus gilae*). *Mitochondrial DNA Part B* **8**, 809–814. https://doi.org/10.1080/23802359.2023.2241658 (2023).
2. Wang, J.-F. et al. Phylogenetic and evolutionary comparison of mitogenomes reveal adaptive radiation of lampriform fishes. *Int. J. Mol. Sci.* **24**, 8756. https://doi.org/10.3390/ijms24108756 (2023).
3. Mthethwa, S., Bester-van der Merwe, A. E. & Roodt-Wilding, R. Addressing the complex phylogenetic relationship of the Gempylidae fishes using mitogenome data. *Ecol. Evolut.* **13**, e10217. https://doi.org/10.1002/ece3.10217 (2023).
4. Nicholls, T. J. & Minczuk, M. In D-loop: 40 years of mitochondrial 7S DNA. *Exp. Gerontol.* **56**, 175–181. https://doi.org/10.1016/j.exger.2014.03.027 (2014).
5. Saccone, C. et al. Evolution of the mitochondrial genetic system: An overview. *Gene* **261**, 153–159. https://doi.org/10.1016/S0378-1119(00)00484-4 (2000).
6. Adams, K. L., Qiu, Y.-L., Stoutemyer, M. & Palmer, J. D. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl. Acad. Sci.* **99**, 9905–9912. https://doi.org/10.1073/pnas.042694899 (2002).
7. Tipton, M. L., Gignoux-Wolfsohn, S., Stonebraker, P. & Chernoff, B. Postglacial recolonization of eastern Blacknose Dace, *Rhinichthys atratulus* (Teleostei: Cyprinidae), through the gateway of New England. *Ecol. Evolut.* **1**, 343–358. https://doi.org/10.1002/ece3.31 (2011).
8. Bernt, M., Braband, A., Schierwater, B. & Stadler, P. F. Genetic aspects of mitochondrial genome evolution. *Mol. Phylogenet. Evolut.* **69**, 328–338. https://doi.org/10.1016/j.ympev.2012.10.020 (2013).
9. Miya, M. & Nishida, M. The mitogenomic contributions to molecular phylogenetics and evolution of fishes: A 15-year retrospect. *Ichthyol. Res.* **62**, 29–71. https://doi.org/10.1007/s10228-014-0440-9 (2015).
10. Hirayama, M. et al. Intraspecific variation in the mitochondrial genome among local populations of Medaka Oryzias latipes. *Gene* **457**, 13–24. https://doi.org/10.1016/j.gene.2010.02.012 (2010).
11. Takada, M. et al. Biogeography and evolution of the *Carassius auratus*-complex in East Asia. *BMC Evolut. Biol.* **10**, 7. https://doi.org/10.1186/1471-2148-10-7 (2010).
12. Osborne, M. J. et al. The complete mitochondrial genomes of three imperiled cyprinid fishes Bonytail (Gila elegans), Rio Grande Silvery Minnow (*Hybognathus amarus*) and Loach Minnow (*Tiaroga cobitis*). *Mitochondrial DNA Part B* **5**, 2368–2370. https://doi.org/10.1080/23802359.2020.1774435 (2020).
13. Robins, J. H. et al. Evolutionary relationships and divergence times among the native rats of Australia. *BMC Evolut. Biol.* **10**, 375. https://doi.org/10.1186/1471-2148-10-375 (2010).
14. Kraczkowski, M. L. & Chernoff, B. Molecular phylogenetics of the Eastern and Western Blacknose Dace, *Rhinichthys atratulus* and *R. obtusus* (Teleostei: Cyprinidae). *Copeia* **2014**, 325–338. https://doi.org/10.1643/CG-14-002 (2014).
15. Schönhuth, S., Vukić, J., Šanda, R., Yang, L. & Mayden, R. L. Phylogenetic relationships and classification of the Holarctic family Leuciscidae (Cypriniformes: Cyprinoidei). *Mol. Phylogenet. Evolut.* **127**, 781–799. https://doi.org/10.1016/j.ympev.2018.06.026 (2018).
16. Schönhuth, S., Shiozawa, D. K., Dowling, T. E. & Mayden, R. L. Molecular systematics of western North American cyprinids (Cypriniformes: Cyprinidae). *Zootaxa* **3586**, 281–303. https://doi.org/10.11646/zootaxa.3586.1.27 (2012).
17. Ramos, E. K. D. S., Freitas, L. & Nery, M. F. The role of selection in the evolution of marine turtles mitogenomes. *Sci. Rep.* **10**, 16953. https://doi.org/10.1038/s41598-020-73874-8 (2020).
18. Miya, M. et al. Major patterns of higher teleostean phylogenies: A new perspective based on 100 complete mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **26**, 121–138. https://doi.org/10.1016/S1055-7903(02)00332-9 (2003).
19. Li, T. et al. Long-branch attraction and the phylogeny of true water bugs (Hemiptera: Nepomorpha) as estimated from mitochondrial genomes. *BMC Evolut. Biol.* **14**, 99. https://doi.org/10.1186/1471-2148-14-99 (2014).
20. Ontano, A. Z. et al. Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions. *Mol. Biol. Evolut.* **38**, 2446–2467. https://doi.org/10.1093/molbev/msab038 (2021).
21. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175. https://doi.org/10.1038/s41592-020-01056-5 (2021).
22. Laetsch, D. & Blaxter, M. BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research* https://doi.org/10.12688/f1000research.12232.1 (2017).
23. Sayers, E. W. et al. Database resources of the national center for biotechnology information. *Nucl. Acids Res.* **50**, D20-d26. https://doi.org/10.1093/nar/gkab1112 (2022).
24. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025 (2020).
25. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191 (2018).
26. Seqtk: Toolkit for processing sequences in FASTA/Q formats v. 1.3 (2012).
27. Camacho, C. et al. BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421. https://doi.org/10.1186/1471-2105-10-421 (2009).
28. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucl. Acids Res.* **33**, D501–D504. https://doi.org/10.1093/nar/gki025 (2005).
29. Jain, C., Rhie, A., Hansen, N. F., Koren, S. & Phillippy, A. M. Long-read mapping to repetitive reference sequences using Winnowmap2. *Nat. Methods* **19**, 705–710. https://doi.org/10.1038/s41592-022-01457-8 (2022).
30. Iwasaki, W. et al. MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol. Evolut.* **30**, 2531–2540. https://doi.org/10.1093/molbev/mst141 (2013).
31. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539. https://doi.org/10.1038/msb.2011.75 (2011).
32. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2023).
33. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw* **4**, 1686. https://doi.org/10.21105/joss.01686 (2019).
34. Paradis, E. & Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528. https://doi.org/10.1093/bioinformatics/bty633 (2019).
35. Schliep, K. P. phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593. https://doi.org/10.1093/bioinformatics/btq706 (2011).

36. data.table: Extension of 'data.frame' v. R package version 1.15.4 (2024).
37. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T.T.-Y. ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evolut.* **8**, 28–36. https://doi.org/10.1111/2041-210X.12628 (2017).

Acknowledgements
We thank the people at Cantata Bio LLC, especially Sierra McWilson and Oli Jin, for producing an excellent quality assembled genome of Rhinichthys atratulus. The authors are very grateful to Professor Masaki Miya, Natural History Museum and Institute, Chiba, Japan. We express our thanks to Laurie Kenney, Jennifer Enxuto, Shana Sperry, Phaedra Durost and Suzanne Bussolari for administrative support. Support to NW was generously provided by Menakka and Essel Bailey Jr. Undergraduate Experience Funds. We are grateful to the Robert K. Schumann Foundation for student research support through the Robert K. Schumann Institute of the Bailey College of the Environment. Grants in Support of Scholarship by Academic Affairs, Wesleyan University, were awarded to BC.

Author contributions
TSE and NW contributed equally to this work. BC and AMA collected the tissue sample. TSE, NW, and BC designed the study. TSE designed the analysis pipeline. NW and TSE analyzed and visualized the data. TSE, NW, SJT, AMA, and BC contributed to the interpretation of the data and provided input on the manuscript, which was prepared by TSE, NW, and BC.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-83134-8.

Correspondence and requests for materials should be addressed to T.S.E.

Reprints and permissions information is available at www.nature.com/reprints.

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