



# First Report of Chromosome-Level Genome Assembly for Flathead Grey Mullet, *Mugil cephalus* (Linnaeus, 1758)

Mudagandur S. Shekhar<sup>1</sup>, Vinaya Kumar Katneni<sup>1\*</sup>, Ashok Kumar Jangam<sup>1</sup>, Karthic Krishnan<sup>1</sup>, Sudheesh K. Prabhudas<sup>1</sup>, Jesudhas Raymond Jani Angel<sup>1</sup>, Krishna Sukumaran<sup>2</sup>, Muniyandi Kailasam<sup>2</sup> and Joykrushna Jena<sup>3</sup>

<sup>1</sup>Nutrition Genetics and Biotechnology Division, Indian Council of Agricultural Research-Central Institute of Brackishwater Aquaculture, Chennai, India, <sup>2</sup>Finfish Culture Division, ICAR-Central Institute of Brackishwater Aquaculture, Chennai, India, <sup>3</sup>Indian Council of Agricultural Research, New Delhi, India

**Keywords:** whole genome, repeat analysis, synteny, grey mullet, PacBio Sequel II, genome annotation

## INTRODUCTION

Amid dwindling production from capture fisheries, aquaculture is an important source of high-quality protein (FAO Food and Agriculture Organization of the United Nations, 2020). In addition, aquaculture generates employment and contributes to eradicating poverty. *Mugil cephalus*, flathead grey mullet, is globally distributed and inhabits inshore seas, estuaries, and brackish water areas (Whitfield et al., 2012). It is of important commercial value to global fisheries and aquaculture with high demand for mullet roe. This species belongs to the family Mugilidae which comprises 26 genera and 80 valid species (Fricke et al., 2022). The adaptability of *M. cephalus* to varied aquatic environments at different life stages and tolerance to a wide range of salinities and temperatures occurring in tropical, subtropical, and temperate coastal waters make it an important cultivatable fish species across the world. The whole-genome sequence information for an aquaculture species has potential applications in genomic selection and selective breeding for sustainable production and improvement of desirable traits, such as disease resistance and growth.

Within the family Mugilidae, a draft genome assembly was first reported for redlip mullet, *Liza haematocheila* (0.74 Gb), which has 1,453 contigs with an N50 length of 3.9 Mb (Liyanage et al., 2019). Later, chromosome-level assembly of this fish was generated with Oxford Nanopore long-read, single-tube long fragment reads (stLFR), and HiC chromatin interaction data, which are 652.91 Mb length in 514 scaffolds with contig and scaffold N50 lengths of 7.21 and 28.01 Mb, respectively (Zhao et al., 2021). For *M. cephalus*, no whole-genome assembly with long-read data is available. Previous reports have attempted genome assembly of this fish at a very low sequence depth using Illumina technology (Dor et al., 2016, 2020). Thus, in the absence of a reference genome for *M. cephalus*, this study aimed to decipher the whole-genome sequence, which will provide baseline information needed to implement genetic improvement programs. The integration of genome information into fisheries and aquaculture management is important to ensure long-term sustainable fishery harvest and aquaculture production. In the present study, a combination of PacBio, Illumina, and Arima Hi-C technologies were applied to construct the genome assembly of *M. cephalus*, an economically important brackish water aquaculture species.

## Value of Data

- The whole-genome sequence assembly generated for *M. cephalus* can be used as a reference genome for the family Mugilidae.
- The high-quality, chromosome-level genome assembly along with the predicted protein sequences would help gain insights into desirable traits through gene expression studies.

## OPEN ACCESS

### Edited by:

Xu Wang,  
Auburn University, United States

### Reviewed by:

Eric M. Hallerman,  
Virginia Tech, United States  
Xiaoli Ma,  
Jiangsu Normal University, China

### \*Correspondence:

Vinaya Kumar Katneni  
Vinaya.Katneni@icar.gov.in

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 02 April 2022

**Accepted:** 06 May 2022

**Published:** 17 June 2022

### Citation:

Shekhar MS, Katneni VK, Jangam AK,  
Krishnan K, Prabhudas SK,  
Jani Angel JR, Sukumaran K,  
Kailasam M and Jena J (2022) First  
Report of Chromosome-Level  
Genome Assembly for Flathead Grey  
Mullet, *Mugil cephalus*  
(Linnaeus, 1758).  
Front. Genet. 13:911446.  
doi: 10.3389/fgene.2022.911446

**TABLE 1** | Summary of *Mugil cephalus* genome assembly and annotation.

Features	Statistics
No. of contigs	848
Contig N50 size (Mbp)	20.15
Contig L50	14
Contig total length (Mbp)	643.89
No. of scaffolds	583
Scaffold N50 size (Mbp)	28.32
Scaffold L50	10
Total scaffold length (Mbp)	644.11
Number of pseudo-chromosomes	24
Total length of pseudo-chromosomes (Mbp)	634.85
GC %	41.9
No. of protein coding genes	27,269

– The whole-genome assembly provides baseline information needed to implement genetic improvement programs for this commercially important species.

## MATERIALS AND METHODS

### Specimen of *M. cephalus*

A specimen of *M. cephalus* maintained at the Muthukadu Experimental Station of ICAR—CIBA (Chennai, India) was used for generating the sequence data required for genome assembly. The species identity of the specimen was confirmed based on the partial sequence of the barcode gene, cytochrome C oxidase I (COI). Briefly, genomic DNA was isolated using the conventional CTAB method (Mirimin and Roodt-Wilding, 2015) from muscle tissue, and its concentration and quality were assessed using a nanodrop 2000C (Thermo Scientific, Waltham, Massachusetts, United States). A partial fragment of the COI gene was amplified using high-fidelity 2X PCR Master Mix (New England Biolabs, Ipswich, Massachusetts, United States) with the primers, F2 - 5'TCGACTAATCACAAAGAC ATCGGCAC3' and R1 - 5'TAGACTTCTGGGTGGCCAAAG AATCA3' (Ward et al., 2005). The amplification conditions were initial denaturation at 98°C for 30 s; 32 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s; followed by a final extension at 72°C for 2 min. The PCR product of 707 bp was gel-extracted using a QIAquick gel extraction kit (Qiagen, Hilden, Germany) and sequenced bidirectionally using an ABI 3730 sequencer (Applied Biosystems, Waltham, Massachusetts, United States). The partial sequence of the COI gene was submitted to GenBank with accession number, MW584357. The sequence of MW584357 was subjected to phylogenetic analysis along with 476 accessions of the genus *Mugil* (Supplementary Table S1) sourced from the Barcode of Life Data systems database<sup>1</sup> and accession of *Chelon labrosus* as an outgroup. Initially, all the 478 sequences were aligned in BioEdit version 7.2.5 (Hall, 1999) to generate a consensus alignment (516 bp) which was used to build a Maximum Likelihood tree in RAxML version 8.2.9 (Stamatakis,

2014) with the GTRGAMMAI model and a random seed value of 12,345. The final tree was visualized using FigTree v1.4.3<sup>2</sup>, which revealed the clustering of sequence MW584357 with other accessions of *M. cephalus* (Supplementary Figure S1).

### PacBio Sequel II Data Generation

Genomic DNA was extracted from the muscle tissue of *M. cephalus* using the blood and cell culture DNA midi kit (Qiagen). DNA quantification was carried out using a Qubit 4.0 fluorometer (ThermoFisher Scientific). Library preparation was performed with the SMRTbell<sup>®</sup> Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, California, United States), and size selection was carried out using BluePippin<sup>™</sup> (Sage Science, Beverly, Massachusetts, United States). The library was sequenced on the PacBio Sequel II platform (Pacific Biosciences) to generate the sequence data. About 257.7 Gb of sequence data in 15,028,480 subreads was generated with a subread N50 of 28,748 bp (Supplementary Table S2).

### Illumina Data Generation

Genomic DNA extracted from the muscle tissue of *M. cephalus* was used for Illumina library preparation with an insert size of 200–300 bp using the NEBNext<sup>®</sup> Ultra<sup>™</sup> II DNA Library Prep Kit for Illumina<sup>®</sup> (New England Biolabs). The PCR products used for the construction of the library were purified with the AMPure XP reagent (Beckman Coulter, Brea, California, United States). The library was checked for size distribution by Agilent 2,100 Bioanalyzer (Agilent Technologies, Santa Clara, California, United States) and sequenced on Illumina NovaSeq 6,000, S4 Flow Cell (2 × 150 bp read length). About 640 million reads/96.1 Gb data were generated with 93% of Q30 bases (Supplementary Table S3). The Illumina paired-end reads were used to assess genome length and also to correct the erroneous bases in genome assembly contigs.

### HiC Data Generation

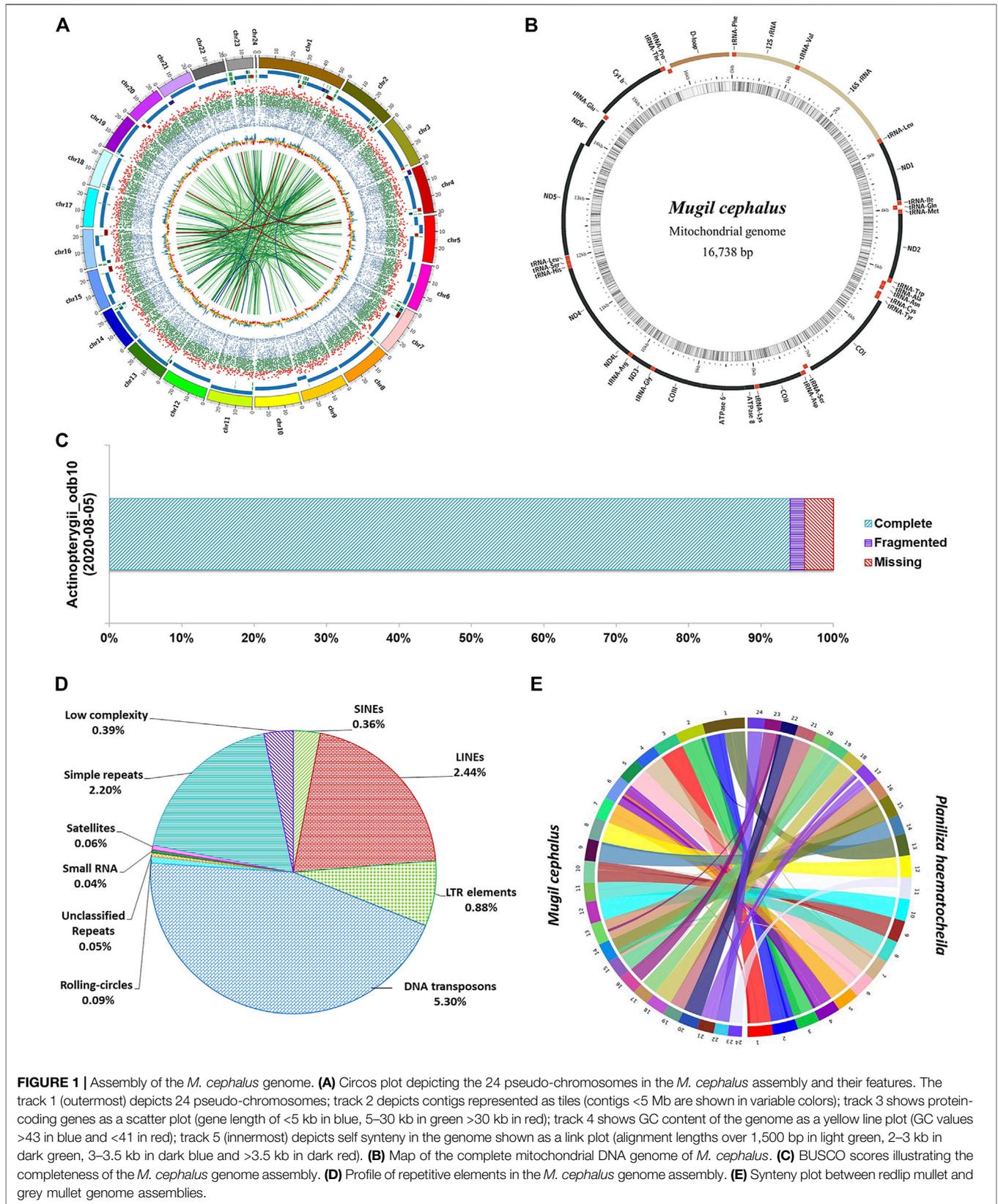
For HiC data generation, tissue crosslinking and proximity ligation were performed using the Phase Genomics Kit (Phase Genomics, Seattle, Washington, United States) followed by Illumina compatible sequencing library preparation. The library was sequenced on the Illumina NovaSeq6000 platform in a 150-bp paired-end mode to generate 181 million paired reads (54.31 Gb), of which 90.74% (49.28 Gb) of bases had a Q30 quality score (Supplementary Table S4). The HiC reads were used in the scaffolding of assembly contigs.

### Genome Length Assessment

The *k*-mer frequency for the grey mullet genome was estimated from the Illumina paired-end reads using Jellyfish 2.2.3 (Marçais and Kingsford, 2011) to count the canonical 21 *k*-mers with the hash size as 20G. The count histogram was later provided as the input to the online tool Genomescope (Vurture et al., 2017) to estimate genome haploid length, heterozygosity, and repeat content. The *k*-mer analysis revealed the genome haploid length to be 594 Mb and genome repeat length as 48 Mb

<sup>1</sup>www.boldsystems.org

<sup>2</sup>http://tree.bio.ed.ac.uk/software/figtree/



(**Supplementary Figure S2**). The genome size of *M. cephalus* was reported to be 857 Mb based on the flow cytometry principle (Raymond et al., 2022).

## Genome Assembly

Genome assembly was performed with the WTDBG2.5 (Ruan and Li, 2020) tool by limiting the usage of PacBio subreads to >30 kb length covering 100 Gb of data. The *de novo* contig-level assembly contained 848 contigs with an N50 length of 20.15 Mb. Polishing of these contigs for base errors and indels with Illumina short reads using POLCA (Zimin and Salzberg, 2020) brought the base consensus quality to 99.99% (**Supplementary Table S5**). Scaffolding was performed on the polished assembly by using juicer and 3D-DNA scripts from the Genome Assembly Cookbook (Dudchenko et al., 2017). Initially, site positions were generated based on the DpnII enzyme used for generating 181 million HiC read pairs. Later, a juicer script was used to generate contact maps in 3D space, and a 3D-DNA script was used to anchor fragments to pseudo-chromosomes which were visualized in Juicebox v1.11.08 (**Supplementary Figure S3**).

At the scaffold level, the assembly is 644 Mb in length in 583 scaffolds with an N50 of 28.32 Mb. The *M. cephalus* karyotype is reported to contain 24 chromosome pairs (Gornung et al., 2001). Accordingly, the longest 24 scaffolds represented 98.56% (634 Mb) of the scaffold-level assembly length and hence are designated as pseudo-chromosomes (**Table 1**). The mitochondrial genome of *Mugil cephalus* was obtained as a single scaffold of 16,738 bp in the assembly. The mitochondrial genome was annotated (**Figure 1B**) using the MitoAnnotator tool<sup>3</sup> (Iwasaki et al., 2013). The genome assembly presented for *M. cephalus* in this study is superior to the assembly reported by Dor et al. (2016) with 480,389 scaffolds and to that reported by Dor et al. (2020) with 4,505 scaffolds in terms of N50 statistics and average scaffold length. The genome assembly reported in this study is of a shorter length than the estimated genome length reported in previous studies (Hinegardner and Rosen, 1972; Dor et al., 2016, 2020; Raymond et al., 2022). The genome completeness for *M. cephalus* assembly was assessed to be 96% complete (**Figure 1C**) using BUSCO v5.2.2 (Seppey et al., 2019) against the Actinopterygii\_odb10 (2020-08-05) orthologous dataset (Kriventseva et al., 2019).

## Repeat Prediction

Repeat masking was performed by using the Repeat masker v4.0.9<sup>4</sup> available in the genome analysis module of Omicsbox v2.0.36 (Bioinformatics, 2019). The RMBlast search was performed for 24 pseudo-chromosomes with a total of 634,849,760 bp against the Eukaryota subset of the RepBase Repeat Masker Edition-20181026<sup>5</sup>. The interspersed, low-complexity, and simple-sequence repeats were masked to obtain the complete repeat profile of *M. cephalus* genome assembly. The total repetitive elements in *M. cephalus* assembly constitute 11.72% (74,376,509 bp) of assembly length as shown in **Supplementary Table S6**. DNA transposons (5.3%,

33,664,198 bp) and the satellite repeats (0.06%, 362,353 bp) contributed the highest and the lowest, respectively, to the total repeat content of the genome (**Figure 1D**). The LINES were the major contributor for retroelements, followed by LTR elements and SINEs accounting for 2.44, 0.88, and 0.36%, respectively. Simple-sequence repeats and low-complexity repeats covered 2.2% (13,966,215 bp) and 0.39% (2,504,867 bp) of the genome, respectively. The MISA v1.0 tool was used to generate the simple sequence repeat profile containing a total of 491,676 SSR elements in assembly (**Supplementary Figure S4**).

## Gene Prediction

The pseudo-chromosomes soft-masked for interspersed sequence repeats avoiding the low-complexity and simple-sequence repeats were used for gene prediction. Structural annotation of the *Mugil cephalus* genome was performed by following the approach as described previously (Katneni et al., 2022), with minor modifications. A combination of *ab initio* and evidence-based approaches was used to obtain the final protein-coding gene set. Briefly, *ab initio* prediction was performed using Augustus v3.3.3 (Stanke et al., 2006) and GeneMark-ES v4.59 (Lomsadze et al., 2005) on the repeat-masked genome. The Illumina paired-end RNA-seq data of fish tissues and developmental stages accessed from GenBank (SRR13039561-66; SRR15243984-88 and SRR16311547-49) were mapped to the genome assembly using Hisat2 v2.2.0 (Kim et al., 2019) aligner. Then, StringTie v2.1.4 (Pertea et al., 2015) was used on this sorted alignment file to assemble the transcripts which were processed in TransDecoder v5.5.0<sup>6</sup> to obtain the Open-Reading Frames (ORFs). Similarly, PacBio IsoSequence reads (NCBI Bioproject: PRJNA675305) were aligned to the genome assembly using GMAP v2020-06-30 (Wu and Watanabe, 2005) to generate valid gene structures using PASA v2.4.1 (Haas et al., 2003) and then finally to generate genome coordinates using TransDecoder v5.5.0<sup>6</sup>. Furthermore, evidence was gathered from the proteins of related species (*Archocentrus centrarchus*, *Astatotilapia calliptera*, *Gambusia affinis*, *Melanotaenia boesemani*, *Maylandia zebra*, *Oreochromis niloticus*, *Oryzias melastigma*, *Parambassis ranga*, *Salarias fasciatus*, and *Xiphophorus maculatus* genome assemblies) by aligning them to the *M. cephalus* genome using GenomeThreader v1.7.3 (Gremme, 2012). Finally, all the evidence from *ab initio* gene predictions, Isosequence reads, RNA-seq reads, and protein data of related species were combined using Evidence Modeler (Haas et al., 2008) to obtain the consensus and non-redundant protein-coding gene set containing 30,882 genes. Then, AGAT<sup>7</sup> was used to filter out the incomplete gene models, resulting in the final gene set of 27,269 genes (**Supplementary Table S7**). Noncoding genes in the *M. cephalus* genome were identified by aligning the repeat-masked assembly to the Rfam<sup>8</sup> database using infernal v1.1.4 (Nawrocki and Eddy, 2013). A total of 5,165 non-coding RNAs were identified (**Supplementary Table S8**), out of which 5S\_RNA, tRNA, and histone3 were in abundance.

<sup>3</sup><http://mitofish.aori.u-tokyo.ac.jp/annotation/input.html>

<sup>4</sup><http://www.repeatmasker.org/>

<sup>5</sup><https://www.girinst.org/repbase/>

<sup>6</sup><https://github.com/TransDecoder>

<sup>7</sup><https://github.com/NBISweden/AGAT>

<sup>8</sup><https://rfam.xfam.org/>

## Functional Annotation

Homology-based genome annotation was performed with the blastx tool (Altschul et al., 1990) using the Actinopterygii (txid7898) non-redundant database of GenBank. The protein domains and orthology-based annotations were executed by the Interproscan module and EggNOG mapper module (Huerta-Cepas et al., 2019), as implemented in OmicsBox v2.0.36 (Bioinformatics, 2019), respectively. Then, all the annotations were merged using OmicsBox v2.0.36 (Bioinformatics, 2019) to obtain the gene ontology of the annotated genes and map the enzyme codes. The pathway details of the annotated genes were obtained by mapping against the KEGG database (Kanehisa et al., 2017). Blastx hits were obtained for 22,334 (81.9%) transcripts from which 18,806 (69%) were functionally annotated and 8,113 (29.8%) were mapped with enzyme codes (Supplementary Figure S5). The transferases and hydrolases were the most dominant enzyme classes expressed (Supplementary Figure S6). Gene ontology analysis revealed that the most expressed GO categories were gene expression (Biological processes), metal ion binding (Molecular function), and intracellular membrane-bounded organelle (Cellular components) in *M. cephalus* proteins (Supplementary Figure S7).

## Synteny Analysis Between *Mugil cephalus* and *Planiliza haematocheila*

Comparative genome analysis was performed to characterize genome-wide syntenic regions between redlip mullet (Zhao et al., 2021) and grey mullet (this study) genomes. The 24 chromosome-level scaffolds of the redlip mullet assembly (648.41 Mb) obtained from the China National Gene Bank Database (CNGDB: CNP0001604) were aligned to the 24 pseudo-chromosomes of the grey mullet assembly (634.84 Mb) using SyMAP 4.2 (Soderlund et al., 2006). A total of 93 synteny blocks were observed between the two genomes. Among them, 24 blocks of the grey mullet genome were found to align against 25 blocks of the redlip mullet genome with block lengths greater than 10 Mb (Figure 1E; Supplementary Table S9).

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

## REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2
- Bioinformatics, B. (2019). OmicsBox-Bioinformatics Made Easy. Available at: <https://www.biobam.com/omicsbox/>.
- Dor, L., Shirak, A., Curzon, A. Y., Rosenfeld, H., Ashkenazi, I. M., Nixon, O., et al. (2020). Preferential Mapping of Sex-Biased Differentially-Expressed Genes of Larvae to the Sex-Determining Region of Flathead Grey Mullet (*Mugil cephalus*). *Front. Genet.* 11, 839. doi:10.3389/fgene.2020.00839
- Dor, L., Shirak, A., Rosenfeld, H., Ashkenazi, I. M., Band, M. R., Korol, A., et al. (2016). Identification of the Sex-Determining Region in Flathead Grey Mullet (*Mugil cephalus*). *Anim. Genet.* 47, 698–707. doi:10.1111/age.12486

accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, PRJNA675305; <https://figshare.com/>, <https://doi.org/10.6084/m9.figshare.19499054.v1>.

## ETHICS STATEMENT

The animal study was reviewed and approved by the Institute Animal Ethics Committee (CIBA/IAEC/2021-25) of ICAR—Central Institute of Brackishwater Aquaculture, Chennai, India.

## AUTHOR CONTRIBUTIONS

MS and JoJ, designed and conducted the study. VK, AJ, KK, and SP performed the sequencing analysis and genome assembly. JeJ, KS, and MK provided the samples and conducted phylogenetic analysis. MS wrote the manuscript with inputs from all authors. All the authors read and approved the final version of the manuscript.

## FUNDING

This work was supported by ICAR-Consortia Research Platform on Genomics, Indian Council of Agricultural Research, New Delhi, India.

## ACKNOWLEDGMENTS

The authors acknowledge the Director, ICAR-CIBA for providing the infrastructure and support required to conduct the study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.911446/full#supplementary-material>

- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). De Novo assembly of the *Aedes aegypti* Genome Using Hi-C Yields Chromosome-Length Scaffolds. *Science* 356, 92–95. doi:10.1126/science.aal3327
- FAO Food and Agriculture Organization of the United Nations (2020). The State of World Fisheries and Aquaculture 2020. Available at: <https://www.fao.org/documents/card/en/c/ca9229en/>.
- Fricke, R., Eschmeyer, W. N., and Fong, J. D. (2022). Eschmeyer's Catalog of Fishes: Genera/Species by Family/Subfamily. Available at: <http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp> (accessed on 01 Mar, 2022).
- Gornung, E., Cordisco, C., Rossi, A., Innocenti, D. S., Crosetti, D., and Sola, L. (2001). Chromosomal Evolution in Mugilidae: Karyotype Characterization of *Liza saliens* and Comparative Localization of Major

- and Minor Ribosomal Genes in the Six Mediterranean Mullet. *Mar. Biol.* 139, 55–60.
- Gremme, G. (2012). *Computational Gene Structure Prediction*. Hamburg: Universität Hamburg. PhD thesis.
- Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Hannick, L. I., et al. (2003). Improving the *Arabidopsis* Genome Annotation Using Maximal Transcript Alignment Assemblies. *Nucleic Acids Res.* 31, 5654–5666. doi:10.1093/nar/gkg770
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated Eukaryotic Gene Structure Annotation Using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7–R22. doi:10.1186/gb-2008-9-1-r7
- Hall, T. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hinegardner, R., and Rosen, D. E. (1972). Cellular DNA Content and the Evolution of Teleostean Fishes. *Am. Nat.* 106, 621–644. doi:10.1086/282801
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. *Nucleic Acids Res.* 47, D309–D314. doi:10.1093/nar/gky1085
- Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., et al. (2013). MitoFish and MitoAnnotator: A Mitochondrial Genome Database of Fish with an Accurate and Automatic Annotation Pipeline. *Mol. Biol. Evol.* 30, 2531–2540. doi:10.1093/molbev/mst141
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* 45, D353–D361. doi:10.1093/nar/gkw1092
- Katneni, V. K., Shekhar, M. S., Jangam, A. K., Krishnan, K., Prabhudas, S. K., Kaikkolante, N., et al. (2022). A Superior Contiguous Whole Genome Assembly for Shrimp (*Penaeus Indicus*). *Front. Mar. Sci.* 8, 8354. doi:10.3389/fmars.2021.808354
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* 37, 907–915. doi:10.1038/s41587-019-0201-4
- Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., et al. (2019). OrthoDB V10: Sampling the Diversity of Animal, Plant, Fungal, Protist, Bacterial and Viral Genomes for Evolutionary and Functional Annotations of Orthologs. *Nucleic Acids Res.* 47, D807–D811. doi:10.1093/nar/gky1053
- Liyanage, D. S., Oh, M., Omeka, W. K. M., Wan, Q., Jin, C. N., Shin, G. H., et al. (2019). First Draft Genome Assembly of Redlip Mullet (*Liza Haematocheila*) from Family Mugilidae. *Front. Genet.* 10, 1246. doi:10.3389/fgene.2019.01246
- Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., and Borodovsky, M. (2005). Gene Identification in Novel Eukaryotic Genomes by Self-Training Algorithm. *Nucleic Acids Res.* 33, 6494–6506. doi:10.1093/nar/gki937
- Marçais, G., and Kingsford, C. (2011). A Fast, Lock-free Approach for Efficient Parallel Counting of Occurrences of *K*-Mers. *Bioinformatics* 27, 764
- Mirimin, L., and Roodt-Wilding, R. (2015). Testing and Validating a Modified CTAB DNA Extraction Method to Enable Molecular Parentage Analysis of Fertilized Eggs and Larvae of an Emerging South African Aquaculture Species, the Dusky Kob *Argyrosomus Japonicus*. *J. Fish. Biol.* 86, 1218–1223. doi:10.1111/jfb.12639
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold Faster RNA Homology Searches. *Bioinformatics* 29, 2933–2935. doi:10.1093/bioinformatics/btt509
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122
- Raymond, J. A. J., Shekhar, M. S., Katneni, V. K., Jangam, A. K., Prabhudas, S. K., Kaikkolante, N., et al. (2022). Comparative Genome Size Estimation of Different Life Stages of Grey Mullet, *Mugil cephalus* Linnaeus, 1758 by Flow Cytometry. *Aquac. Res.* 53, 1151–1158. doi:10.1111/are.15639
- Ruan, J., and Li, H. (2020). Fast and Accurate Long-Read Assembly with Wtdbg2. *Nat. Methods* 17, 155–158. doi:10.1038/s41592-019-0669-3
- Seppy, M., Manni, M., and Zdobnov, E. M. (2019). “BUSCO: Assessing Genome Assembly and Annotation Completeness,” in *Gene Prediction*. Editor M. Kollmar (Humana, NY: Springer), 227–245. doi:10.1007/978-1-4939-9173-0\_14
- Soderlund, C., Nelson, W., Shoemaker, A., and Paterson, A. (2006). SyMAP: A System for Discovering and Viewing Syntenic Regions of FPC Maps. *Genome Res.* 16, 1159–1168. doi:10.1101/gr.5396706
- Stamatakis, A. (2014). RAxML Version 8: a Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: *Ab Initio* Prediction of Alternative Transcripts. *Nucleic Acids Res.* 34, W435–W439. doi:10.1093/nar/gkl200
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: Fast Reference-free Genome Profiling from Short Reads. *Bioinformatics* 33, 2202–2204. doi:10.1093/bioinformatics/btx153
- Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., and Hebert, P. D. N. (2005). DNA Barcoding Australia’s Fish Species. *Phil. Trans. R. Soc. B* 360, 1847–1857. doi:10.1098/rstb.2005.1716
- Whitfield, A. K., Panfili, J., and Durand, J.-D. (2012). A Global Review of the Cosmopolitan Flathead Mullet *Mugil cephalus* Linnaeus 1758 (Teleostei: Mugilidae), with Emphasis on the Biology, Genetics, Ecology and Fisheries Aspects of This Apparent Species Complex. *Rev. Fish. Biol. Fish.* 22, 641–681. doi:10.1007/s11160-012-9263-9
- Wu, T. D., and Watanabe, C. K. (2005). GMAP: A Genomic Mapping and Alignment Program for mRNA and EST Sequences. *Bioinformatics* 21, 1859–1875. doi:10.1093/bioinformatics/bti310
- Zhao, N., Guo, H.-B., Jia, L., Guo, H.-B., Jia, L., Deng, Q.-X., et al. (2021). High-quality Chromosome-Level Genome Assembly of Redlip Mullet (Planiliza Haematocheila). *Zool. Res.* 42, 796–799. doi:10.24272/j.issn.2095-8137.2021.255
- Zimin, A. V., and Salzberg, S. L. (2020). The Genome Polishing Tool POLCA Makes Fast and Accurate Corrections in Genome Assemblies. *PLoS Comput. Biol.* 16, e1007981–8. doi:10.1371/journal.pcbi.1007981

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Shekhar, Katneni, Jangam, Krishnan, Prabhudas, Jani Angel, Sukumaran, Kailasam and Jena. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.