




The draft genome assembly of the cosmopolitan pelagic fish dolphinfish *Coryphaena hippurus*

Silvia Hinojosa-Alvarez ^{1,†}, Verónica Mendoza-Portillo,^{2,‡} Rocío Alejandra Chavez-Santoscoy ¹,
 Jesús Hernández-Pérez,¹ Andrea Felix-Ceniceros,¹ Erika Magallón-Gayón ¹, Adan Fernando Mar-Silva,²
 Mariel Ochoa-Zavala,³ Píndaro Díaz-Jaimes^{2,*}

¹Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Campus Monterrey, Ave. Eugenio Garza Sada 2501 Sur, Monterrey 64849, México

²Unidad Académica de Ecología y Biodiversidad Acuática, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México, Circuito Interior s/n, Ciudad Universitaria, Ciudad de México, C.P. 04510, México

³Escuela Nacional de Estudios Superiores, Unidad Morelia, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro No. 8701. Col. Ex Hacienda de San José de la Huerta. Morelia, Michoacán C.P. 58190, México

*Corresponding author: Unidad Académica de Ecología y Biodiversidad Acuática, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México, Circuito Interior s/n, Ciudad Universitaria, Ciudad de México, Apdo. Postal 70-305, C.P. 04510, México. Email: pindaro@cmarl.unam.mx

†These two authors contributed equally to this work and share co-first authorship.

For the first time, the complete genome assembly of the dolphinfish (*Coryphaena hippurus*), a tropical cosmopolitan species with commercial fishing importance was sequenced. Using a combination of Illumina and Nanopore sequencing technologies, a draft genome of 497.8 Mb was assembled into 6,044 contigs, with an N50 of 200.9 kb and a BUSCO genome completeness score of 89%. This high-quality genome assembly provides a valuable resource to study adaptive evolutionary processes and supports conservation and management strategies for this ecologically and economically significant species.

Keywords: Coryphaenidae; dolphinfish; nanopore; comparative genomics; genome assembly; de novo assembly

Introduction

The dolphinfish (*Coryphaena hippurus*, family Coryphaenidae, order Carangiformes) is a widely distributed pelagic species found in tropical and subtropical waters globally (Farrell *et al.* 2014). Known for its rapid growth rate, short lifespan, and high fecundity, this species plays a critical role in marine food webs, acting as both a predator and prey (Oxenford 1999). Additionally, *C. hippurus* supports significant commercial and recreational fisheries, making it economically valuable and an important target for sustainable management practices (Oxenford 1999). Its broad habitat range, spanning multiple environmental gradients, and high dispersal capacity make it an excellent model for studying the genetic mechanisms of adaptation and population differentiation across diverse oceanic conditions.

Sequencing the genome of *C. hippurus* is crucial to understanding the genetic basis of its adaptations to environmental variability, including temperature, salinity, and food availability. Such insights are especially important given the increasing impact of climate change and commercial exploitation on marine biodiversity (Reiss *et al.* 2009; Pinsky and Palumbi 2014). A reference genome provides a foundational resource for identifying adaptive variants that contribute to resilience in changing environments, as well as for monitoring population health and genetic diversity over time. Furthermore, genomic data from *C. hippurus* can facilitate comparative studies within the order Carangiformes and across other ecologically and economically important fish species (Manel and Holderegger 2013; Riginos and Liggins 2013).

This study aims to produce the first genome assembly for *C. hippurus* using a combination of short-read and long-read sequencing technologies to capture the full genetic complexity of the species. Previous studies have shown that combining these technologies can enhance the quality of reference genomes for various fish species, providing comprehensive tools for population and conservation genomics (Austin *et al.* 2017; Tan *et al.* 2018; Johnson *et al.* 2020; Arias *et al.* 2021). With <1% of actinopterygian species having annotated genomes and limited genomic resources for Carangiformes, this genome will be an invaluable addition to existing resources, supporting both evolutionary studies and conservation efforts (Fan *et al.* 2020).

The genome assembly for *C. hippurus*, will enable researchers to explore genetic, ecological, and evolutionary issues central to understanding adaptation in marine species. This resource will serve also as a model for studying the effects of natural and anthropogenic pressures on marine fish populations, contributing to the sustainable management and conservation of marine biodiversity.

Methods

Sample collection

Specimens of *C. hippurus* were collected from Mazatlan, Mexico through commercial fisheries. A healthy adult fish was selected for genomic analysis, kept in dry ice while transporting it to the laboratory and then preserved at −80°C until further processing.

Table 1. Genomes comparison of closely related species to *C. hippurus*.

Species	Complete BUSCOs	Single copy BUSCOs	Duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	N50 contigs
<i>C. hippurus</i> ^a	88.9%	87.9%	1.0%	2.1%	9.0%	128.5 KB
<i>P. fluviatilis</i> ^b	97.7%	96.9%	0.8%	0.7%	1.6%	4 MB
GCA_010015445.1 (Ozerov et al. 2018)						
<i>P. flavescens</i> ^b	3592	3556	36	13	35	4 MB
GCA_004354835.1 (Feron et al. 2020)						
<i>S. aureovittata</i> ^b	97.7%	96.9%	0.8%	0.7%	1.6%	17 MB
GCA_021018895.1 (Li et al. 2022a)						
<i>S. dorsalis</i> ^c	84.0%	83.2%	0.8%	6.0%	10.0%	11 KB
GCA_002814215.1 (Li et al. 2022b)						
<i>L. japonicus</i> ^c	94.5%	90.1%	4.4%	2.5%	3.0%	285 KB
GCA_033238685.1 (Hashiguchi et al. 2024)						
<i>L. calcarifer</i> ^b	98.7%	95.3%	3.4%	0.4%	0.9%	1 MB
GCA_001640805.2 (Vij et al. 2016)						
<i>C. melampygus</i> ^c	97.8%	89.9%	7.9%	0.7%	1.5%	1 MB
GCA_019059645.1 (Pickett et al. 2021)						

^a This study ^breference genome (chromosome level) ^cscaffold.

DNA extraction

Total genomic DNA was extracted from fin tissue using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol with slight modifications. Tissue samples were re-suspended in 180 µL of ATL buffer and 20 µL of proteinase K. The homogenate was incubated at 56°C for 24 h with intermittent vortex. After digestion, 200 µL of AL buffer was added, mixed, and incubated at room temperature for 10 min. The DNA was purified using DNeasy spin columns and eluted twice in 100 µL of AE buffer at 1000×g for 1 min and 8000×g for another 1 min. DNA concentration was assessed using a Qubit dsDNA HS quantification assay kit (Thermo Fisher Scientific), purity was evaluated using a NanoDrop One (Thermo Fisher Scientific), and DNA integrity was determined through agarose gel electrophoresis. DNA samples with a concentration >100 ng/µL and that showed absence of DNA degradation in agarose gel electrophoresis were selected for sequencing.

Illumina short-read sequencing

A genomic DNA library for Illumina sequencing was prepared using the Illumina DNA Library Prep Kit (Illumina). The library was prepared following the manufacturer's protocol. The library was quantified using a Qubit dsDNA HS quantification assay kit, and its size distribution was verified through capillary electrophoresis using a S2 DNA standard cartridge in a Qsep400 (BioOptic). Sequencing was performed on an Illumina NovaSeq 6000 platform using a 2 × 150 bp paired-end configuration.

Nanopore long-read sequencing

Twenty-four genomic DNA libraries for Nanopore sequencing were prepared from the same genomic DNA aliquot using the Rapid Sequencing Kit V14 (Oxford Nanopore Technologies). 200 ng of high-quality DNA were used to prepare each library according to manufacturer's instructions. Final libraries were quantified using a Qubit dsDNA HS quantification assay kit. Sequencing was performed in a MinION Mk1C device (Oxford Nanopore Technologies) using one MinION flow cell per 6 libraries. Each sequencing run was performed using a new flow cell; 6 MinION flow cells were used to sequence all 36 libraries.

Data quality control

Illumina short-read data quality from RNA and DNA was assessed using FastQC v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and reads were trimmed with Trimmomatic v0.39 (Bolger et al. 2014) to remove adapter sequences and low-quality bases. Nanopore long-read data quality was evaluated using Guppy v5.0.14 (https://community.nanoporetech.com/docs/prepare/library_prep_protocols/Guppy-protocol/v/gpb_2003_v1_revax_14dec2018/guppy-software-overview) for base-calling and NanoPlot v1.36.0 (<https://github.com/wdecoster/NanoPlot>) for read length and quality metrics. Long-read quality was checked with LongQC (Fukasawa et al. 2020), and remaining adapters were trimmed with Porechop (Bonenfant et al. 2023).

Genome assembly

The initial assembly was performed with long reads to generate a draft genome assembly with Flye v 2.9.3 (Kolmogorov et al. 2019) using default parameters. The draft assembly was polished 3 rounds using Illumina short-read data with Pilon v1.24 (Walker et al. 2014) to correct errors, both programs were run on the Omics Box suite (BioBam Bioinformatics, <https://www.biobam.com/omicsbox>) with default parameters.

Assembly evaluation

The quality and completeness of the final genome assembly were evaluated using the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool v5.5.0 (Manni et al. 2021) with the Actinopterygii_odb10 database for genome completeness assessment run on the Omics Box suite under predetermined parameters (<https://www.biobam.com/omicsbox>). Seven genomes of phylogenetic related species *Perca fluviatilis*, *P. flavescens*, *Seriola aureovittata*, *S. dorsalis*, *Lates calcarifer*, *L. japonicus*, and *Caranx melampygus* were downloaded from NCBI (Table 1 and Supplementary Table 1) to compare BUSCO statistics. Assembly metrics, including N50, contig length, and the number of contigs, were analyzed using QUAST v5.2.0 (Gurevich et al. 2013) run under default parameters on the Omics Box suite (BioBam Bioinformatics, <https://www.biobam.com/omicsbox>).

Gene annotation

To improve the gene prediction RNA-Seq data was obtained from dolphinfish tissue control samples (monoil exposed) of BioProject

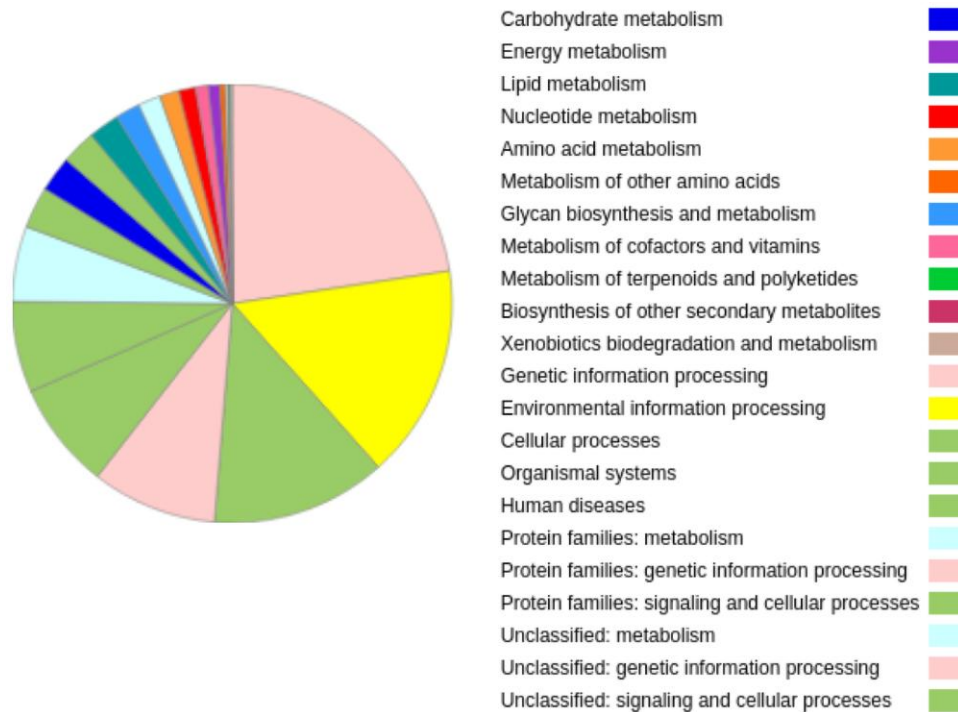


Fig. 1. Functional classification of biological pathways in the *C. hippurus* transcriptome.

Table 2. Summary of repetitive elements in the *C. hippurus* genome.

	Number of elements ^a	Length occupied	Percentage of sequence
Retroelements	26,987	3,895,527 bp	0.78%
SINES:	3127	316,648 bp	0.06%
Penelope:	0	0 bp	0.00%
LINES:	8078	1,590,131 bp	0.32%
CRE/SLACS	0	0 bp	0.00%
L2/CR1/Rex	3883	574,723 bp	0.12%
R1/LOA/Jockey	157	20,673 bp	0.00%
R2/R4/NeSL	411	196,148 bp	0.04%
RTE/Bov-B	1208	442,619 bp	0.09%
L1/CIN4	1604	289,200 bp	0.06%
LTR elements:	15,782	1,988,748 bp	0.40%
BEL/Pao	806	325,022 bp	0.07%
Ty1/Copia	14	6115 bp	0.00%
Gypsy/DIRS1	7075	916,350 bp	0.18%
Retroviral	4838	493,398 bp	0.10%
DNA transposons	62,971	7,317,893 bp	1.47%
hobo-Activator	15,409	1,390,496 bp	0.28%
Tc1-IS630-Pogo	11,908	3,291,932bp	0.66%
En-Spm	0	0 bp	0.00%
MULE-MuDR	851	48,507 bp	0.01%
PiggyBac	767	113,574 bp	0.02%
Tourist/Harbinger	1757	137,063 bp	0.03%
Other (Mirage P-element Transib)	1014	108,422 bp	0.02%
Rolling-circles	772	48,901 bp	0.01%
Unclassified:	173	48,350 bp	0.01%
Total interspersed repeats:		11,261,770bp	2.26%
Small RNA:	1146	100,962 bp	0.02%
Satellites:	619	58,851 bp	0.01%
Simple repeats:	310,761	14,556,967 bp	2.93%
Low complexity:	36,791	2,061,291 bp	0.41%

For each category, the number of elements detected, the total length occupied (in base pairs), and the percentage of the sequence length are provided.

PRJNA763937 (Schlenker et al. 2022), all the control samples were concatenated and then assembled with Trinity v2.15.1 with the following command `Trinity --seqType fq --single/user/dir/*.fastq.gz --CPU 30 --max_memory 150G --output/user/dir/out` (Grabherr et al. 2011). Ab initio prediction of genes was done using the RepeatMasker, (Smit et al. 2013) AUGUSTUS (Stanke et al. 2006) and SNAP (Korf 2004). Final annotation was performed with the assembled transcripts, the vertebrate proteins database, and the ab initio predicated genes using the MAKER pipeline v3.01.03 (<https://hpc.nih.gov/apps/MAKER.html>). Functional annotation of predicted genes was performed using BLAST v2.11.0 against the UniProt and NCBI databases, and InterProScan v5.53-87.0 (<https://www.ebi.ac.uk/interpro/search/sequence/>) for domain identification run the Omics Box suite (BioBam Bioinformatics, <https://www.biobam.com/omicsbox>) using defaults values. A total of 4 rounds of maker were evaluated. The scripts used for MAKER pipeline can be found in https://github.com/Tec-BASE-Laboratory/MAKER_Chippurus.

Comparative genomics

Protein sequences from the same 7 species were retrieved from the NCBI database. Additionally, protein sequences predicted by AUGUSTUS from the draft genome of *C. hippurus* were incorporated into the analysis. Comparative genomic analysis was performed using OrthoFinder v2.5.4 (Emms and Kelly 2015) to identify orthologous gene families across the species. In addition, OrthoVenn3 web page (<https://orthovenn3.bioinfotoolkits.net/>) was used to construct the phylogenetic tree using single-copy orthologues and produce the graphics for orthologous cluster comparison.

Results and discussion

Genome assembly and quality assessment

The final genome assembly comprised 497.8 mega base pairs (Mb), with an N50 of 200.9 kb (Table 1). The largest contig measured

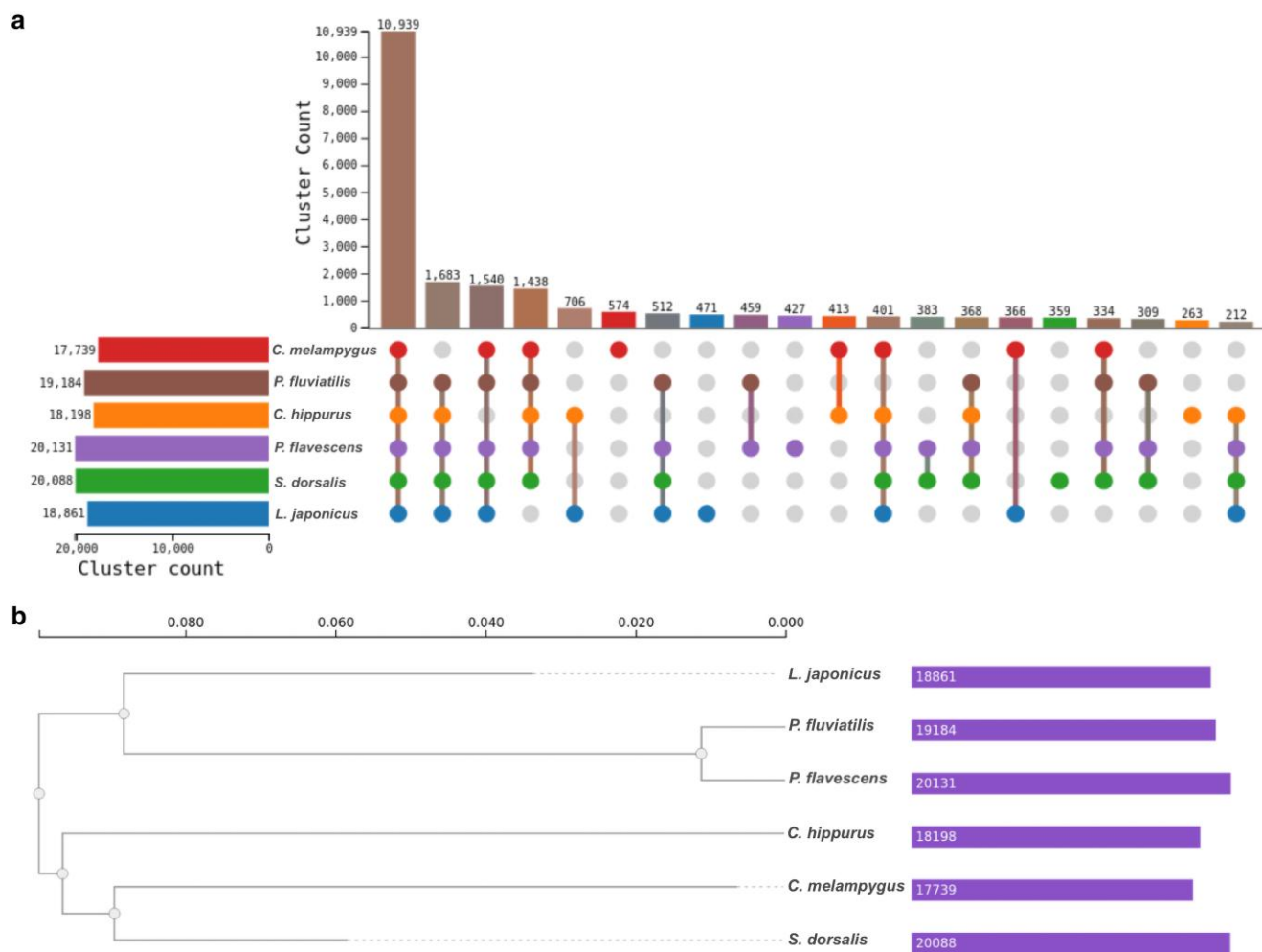


Fig. 2. Comparative genomics of *C. hippurus* with related species. a) This figure shows the identification of orthologous gene clusters across 7 species. The number of shared and unique gene clusters is visualized. b) Phylogenetic tree constructed using single-copy orthologues, providing insights into the evolutionary relationships of *C. hippurus* within Carangiformes.

590,344 bp, while the shortest contig was 5,963 bp. The estimated genome size produced with findGSE (Sun et al. 2018) was 534,132 Mb. BUSCO v5.5.0 analysis using the actinopterygii_odb10 core gene set revealed that 89.0% of the reference dataset sequences had a complete ortholog, and 88.0% of the genes were complete single-copy genes. While the draft genome of *C. hippurus* shows feasibility for refinement compared with other genome assemblies of closely related species, it provides a robust and reliable resource for genomic studies.

Genome annotation

Transcriptome assembly

The BUSCO analysis of the transcriptome assembly indicated that 57.09% of orthologs were present as single copies, while 10% were identified as duplicated, which could reflect biological duplications or technical redundancies, such as splice variants. This analysis highlights the completeness of the transcriptome and its utility in annotating the genome. These results are consistent with the proteins obtained from AUGUSTUS on the assembled transcriptome revealed pathways primarily related to metabolism, environmental processing, and cellular processing (Fig. 1).

Annotation

The results obtained with Repeat Masker (Table 2) showed that retroelements comprise 26,987 elements spanning 3,895,527 bp,

accounting for 0.78% of the total sequence. This category includes subtypes like short interspersed nuclear elements (SINES), interspersed nuclear elements (LINEs), and long terminal repeat elements (LTR), with LINEs being the most prevalent, occupying 1,590,131 bp (0.32%). DNA transposons are more abundant, with 62,971 elements covering 7,317,893 bp, representing 1.47% of the sequence. Additionally, simple repeats are notably prominent, consisting of 310,761 elements and covering 14,556,967 bp, which corresponds to 2.93% of the sequence. The diversity and extent of repetitive sequences provide a clear overview of the genomic composition in terms of repeat content. These results are congruent with those reported for other marine fishes such as King Angelfish (*Holacanthus passer*, e.g. 1.36% of DNA transposons, 2.14% of simple repeats, Gatins et al. 2024) but low in LINEs in comparison with other Carangiforms, for example, Carangidae family members, (Coryphaenidae's sister family) such as *C. melampyus*, which shows values of 2.3% for LINEs and 5.4% for DNA transposons (Pickett et al. 2021), or *Seriola* species (*S. lalandi* 3.85%, Li et al. 2022a; *S. aureovittata* 4.71% of LINEs, Li et al. 2022b).

Comparative genomics

After completing 4 annotation rounds with MAKER, it was observed that round 1 and round 4 showed more gene models 34,129 and 35,670 respectively also higher annotation edit

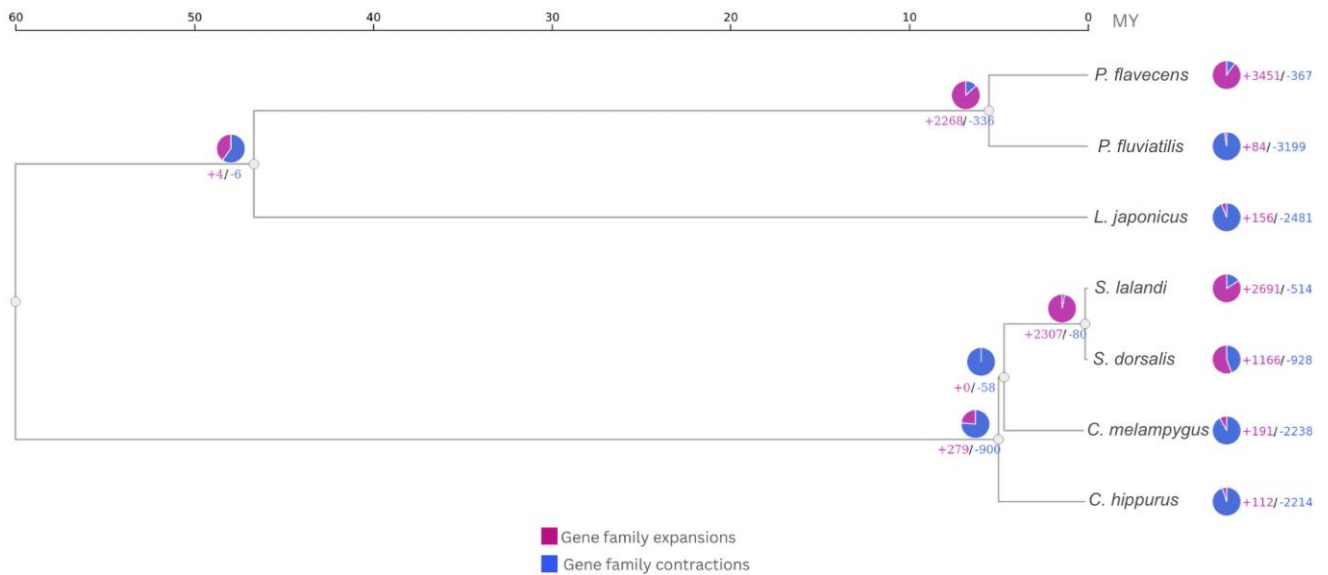


Fig. 3. Gene family expansions (pink) and contractions (blue) in *C. hippurus* and related species. The figure represents gene family dynamics, showing expansions and contractions across species.

Table 3. Enriched pathway in single copy unique orthologues of *C. hippurus*.

GO-ID	Count	Pathway	GO-term	P-value
GO:0016021	2	Integral component of membrane	Cellular component	1.23E-41
GO:0003964	20	RNA-directed DNA polymerase activity	Molecular function	9.01E-26
GO:0006355	2	Regulation of transcription, DNA-templated	Biological process	5.80E-24
GO:0016055	3	Wnt signaling pathway	Biological process	2.71E-19
GO:0007283	2	Spermatogenesis	Biological process	2.48E-16
GO:0016032	2	Viral process	Biological process	1.57E-15
GO:0016192	2	Vesicle-mediated transport	Biological process	7.67E-15
GO:0046872	3	Metal ion binding	Molecular function	4.75E-14
GO:0016567	2	Protein ubiquitination	Biological process	5.87E-11
GO:0032197	8	Transposition, RNA-mediated	Biological process	1.80E-09
GO:0051056	2	Regulation of small GTPase mediated signal transduction	Biological process	0.006397882811536
GO:0006334	2	Nucleosome assembly	Biological process	0.021287316663045

distance with 98 and 96% confidence in the gene models (Supplementary Table 1). Since round 4 had more gene models and good confidence in the gene models this was used as the final annotation.

Figure 2a visualizes the intersections of orthologous gene clusters across the 7 species, with *P. flaveces* having the highest count of 20,131 clusters and *L. japonicus* the lowest with 18,861 clusters. The main bar chart at the top shows the size of the intersections between these species' gene clusters. The largest intersection, containing 10,939 clusters, is shared among all 7 species, indicating a significant core set of orthologous genes. Other notable intersections include smaller subsets shared among different combinations of species, as indicated by the connected dots below the bar chart. For instance, there are 1,683 clusters unique to *P. fluviatilis* and *C. melampygus*, and 1,540 clusters unique to *C. hippurus* and *P. flaveces*. The phylogenetic tree constructed with single-copy orthologues (Fig. 2b) is consistent with the previously described for Carangiformes (Glass et al. 2023).

Gene expansions and contractions are evolutionary processes where gene families either increase or decrease in size. These changes can lead to adaptations, with expansions often providing new functions or enhancing existing ones, while contractions may reflect the loss of redundant or nonessential functions as species evolve and adapt to new environments. In Fig. 3, *P. flaveces*

exhibits a significant number of gene family expansions (+3,451) compared with contractions (−367), while *P. fluviatilis* shows fewer expansions (+84) and a substantial number of contractions (−3,199). *L. japonicus* and *C. melampygus* both display moderate expansions (+156 and +191, respectively) with a higher number of contractions (−2,481 and −2,238, respectively). *S. lalandi* and *S. dorsalis* exhibit high and moderate gene family expansions (+2,691 and +1,166, respectively) and fewer contractions (−514 and −928, respectively). Meanwhile, *C. hippurus* shows a more balanced pattern with +112 expansions and −2,214 contractions. The internal nodes of the tree reflect the evolutionary history of these species, with varying degrees of expansions and contractions, such as +2,268/−336 at 1 node and +2,307/−80 at another. Overall, the analysis indicates a trend where gene family contractions often outnumber expansions, particularly in species like *P. fluviatilis*, *L. japonicus*, *C. melampygus*, and *C. hippurus*, while *P. flaveces* and *S. lalandi* show more pronounced gene family expansions. It is worth mentioning that *C. hippurus* genome is a first version and these missing regions may contribute to false gene contractions, highlighting the need for caution in interpreting gene family dynamics. Despite these missing regions, this genome provides a valuable resource for studying the genetic basis of adaptation, population structure, and conservation in a species of significant ecological and commercial importance.

Coryphaena hippurus presented 263 clusters unique to the species, for these gene clusters, gene ontology (GO, Ashburner et al. 2000) enrichment was calculated, and the enriched pathways are mainly related to biological processes (Table 3). As an example, Wnt signaling pathways are involved in the development of tissues and organs in embryos, and they are controlled by an interplay of several signaling pathways that crosstalk to provide positional information and induce cell fate specification (Huelsenken and Birchmeier 2001).

Conclusions

We present the first genome of the dolphinfish (*C. hippurus*), combining Illumina and Nanopore sequencing technologies. By using a combination of long and short-read sequencing platforms, we successfully assembled a draft genome of a nonmodel species, obtaining 89% of complete orthologous genes, which indicates that the assembly captures a significant portion of the genome's coding and functional elements. This is critical for downstream applications such as comparative genomics, evolutionary studies, and functional annotation. The assembled genome size of 0.49 Gb closely approximates the estimated size of 0.53 Gb, suggesting minimal loss of genomic information during the assembly process. This level of completeness is sufficient to provide a reliable resource to understand the molecular mechanisms underlying the acclimatization and adaptation to both local and global environmental changes of marine species, including dolphinfish, across its distributional range. These insights will be valuable for future studies on evolutionary biology, population genomics, and the design of conservation strategies for *C. hippurus*, and other nonmodel marine species from which scarce information exists, particularly in the face of accelerating climate change.

Data availability

Raw sequencing data and the draft genome assembly are available under the project accession number PRJNA1146410 of GenBank. Additional Supplemental material available at GSA FigShare: <https://doi.org/10.25387/g3.28548704>.

Supplemental material available at G3 online.

Acknowledgments

We gratefully acknowledge the financial support and bioinformatic resources provided by Tecnológico de Monterrey. In particular, we thank the CoreLab Data Science and the CoreLab Genomics for their valuable contributions in enabling the successful completion of this project. V.M.P. thanks CONAHCYT (Consejo Nacional de Humanidades, Ciencia y Tecnología) for granting a posdoctoral scholarship EPM_2023(1).

Funding

This work was supported by the Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) through the Project “Dilucidando los componentes evolutivos de la divergencia genética en el ambiente marino: el dorado *C. hippurus* como modelo de estudio” with grant number CF-2023-G-493. P.D.J. thanks the Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México for internal financial support No. 611 in the acquisition of supplies for the project.

Conflicts of interest

The authors declare no conflicts of interest.

Literature cited

- Arias CF, Dikow RB, McMillan WO, De León LF. 2021. De Novo genome assembly of the electric fish *Brachyhypopomus occidentalis* (hypopomidae, gymnotiformes). *Genome Biol Evol.* 13(10):evab223. <https://doi.org/10.1093/gbe/evab223>.
- Gene Ontology Consortium T, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Michael Cherry J, Davis AP, Dolinski K, Dwight SS, et al. 2000. Gene ontology: tool for the unification of biology NIH Public Access Author Manuscript. *Nat Genet.* 25(1): 25–29. <https://doi.org/10.1038/75556>.
- Austin CM, Tan MH, Harrison KA, Lee YP, Croft LJ, Sunnucks P, Pavlova A, Gan HM. 2017. De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. *Gigascience.* 6(8):1–6. <https://doi.org/10.1093/gigascience/gix063>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics.* 30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bonenfant Q, Noé L, Touzet H. 2023. Porechop_ABI: discovering unknown adapters in Oxford Nanopore technology sequencing reads for downstream trimming. *Bioinform Adv.* 3(1):vbac085. <https://doi.org/10.1093/bioadv/vbac085>.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16(1):157. <https://doi.org/10.1186/s13059-015-0721-2>.
- Fan G, Song Y, Yang L, Huang X, Zhang S, Zhang M, Yang X, Chang Y, Zhang H, Li Y, et al. 2020. Initial data release and announcement of the 10,000 fish genomes project (Fish10K). *Gigascience.* 9(8):giaa080. <https://doi.org/10.1093/gigascience/giaa080>.
- Farrell ER, Boustany AM, Halpin PN, Hammond DL. 2014. Dolphinfish (*Coryphaena hippurus*) distribution in relation to biophysical ocean conditions in the northwest Atlantic. *Fish Res.* 151:177–190. <https://doi.org/10.1016/j.fishres.2013.11.014>.
- Feron R, Zahm M, Cabau C, Klopp C, Roques C, Bouchez O, Eché C, Valière S, Donnadiou C, Haffray P, et al. 2020. Characterization of a Y-specific duplication/insertion of the anti-Müllerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. *Mol Ecol Resour.* 20(2):531–543. <https://doi.org/10.1111/1755-0998.13133>.
- Fukasawa Y, Ermini L, Wang H, Carty K, Cheung MS. 2020. LongQC: a quality control tool for third generation sequencing long read data. *G3 (Bethesda).* 10(4):1193–1196. <https://doi.org/10.1534/G3.119.400864>.
- Gatins R, Arias CF, Sánchez C, Bernardi G, De León LF. 2024. Whole genome assembly and annotation of the King Angelfish (*Holacanthus passer*) gives insight into the evolution of marine fishes of the Tropical Eastern Pacific. *GigaByte.* 2024:1–18. <https://doi.org/10.46471/gigabyte.115>.
- Glass JR, Harrington RC, Cowman PF, Faircloth BC, Near TJ. 2023. Widespread sympatry in a species-rich clade of marine fishes (*Carangoidae*). *Proc R Soc Lond B Biol Sci.* 290(2010):20230657. <https://doi.org/10.1098/rspb.2023.0657>.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652. <https://doi.org/10.1038/nbt.1883>.

- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUASt: quality assessment tool for genome assemblies. *Bioinformatics*. 29(8): 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Hashiguchi Y, Mishina T, Takeshima H, Nakayama K, Tanoue H, Takeshita N, Takahashi H. 2024. Draft genome of akame (*Lates japonicus*) reveals possible genetic mechanisms for long-term persistence and adaptive evolution with low genetic diversity. *Genome Biol Evol*. 16(8):evae174. <https://doi.org/10.1093/GBE/EVAE174>.
- Huelsken J, Birchmeier W. 2001. New aspects of Wnt signaling pathways in higher vertebrates. *Curr Opin Genet Dev*. 11(5):547–553. [https://doi.org/10.1016/S0959-437X\(00\)00231-8](https://doi.org/10.1016/S0959-437X(00)00231-8).
- Johnson LK, Sahasrabudhe R, Gill JA, Roach JL, Froenicke L, Brown CT, Whitehead A. 2020. Draft genome assemblies using sequencing reads from Oxford Nanopore Technology and Illumina platforms for four species of North American *Fundulus* killifish. *Gigascience*. 9(6):giaa067. <https://doi.org/10.1093/gigascience/giaa067>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 37(5): 540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics*. 5(1):59. <https://doi.org/10.1186/1471-2105-5-59>.
- Li S, Liu K, Cui A, Hao X, Wang B, Wang HY, Jiang Y, Wang Q, Feng B, Xu Y, et al. 2022b. A chromosome-level genome assembly of yellowtail kingfish (*Seriola lalandi*). *Front Genet*. 12:825742. <https://doi.org/10.3389/FGENE.2021.825742/BIBTEX>.
- Li M, Xu X, Liu S, Fan G, Zhou Q, Chen S. 2022a. The chromosome-level genome assembly of the Japanese yellowtail jack *Seriola aurovittata* provides insights into genome evolution and efficient oxygen transport. *Mol Ecol Resour*. 22(7):2701–2712. <https://doi.org/10.1111/1755-0998.13648>.
- Manel S, Holderegger R. 2013. Ten years of landscape genetics. *Trends Ecol Evol*. 28(10):614–621. <https://doi.org/10.1016/j.tree.2013.05.012>.
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 38(10): 4647–4654. <https://doi.org/10.1093/molbev/msab199>.
- Oxenford HA. 1999. Biology of the dolphinfish (*Coryphaena hippurus*) in the western central Atlantic: a review. *Sci Mar*. 63(3–4): 277–301. <https://doi.org/10.3989/SCIMAR.1999.63N3-4303>.
- Ozerov MY, Ahmad F, Gross R, Pukk L, Kahar S, Kisand V, Vasemägi A. 2018. Highly continuous genome assembly of Eurasian perch (*Perca fluviatilis*) using linked-read sequencing. *G3 (Bethesda)* 8(12):3737–3743. <https://doi.org/10.1534/G3.118.200768>.
- Pickett BD, Glass JR, Ridge PG, Kauwe JSK. 2021. De novo genome assembly of the marine teleost, bluefin trevally (*Caranx melampygus*). *G3 (Bethesda)* 11(10):jkab229. <https://doi.org/10.1093/G3JOURNAL/JKAB229>.
- Pinsky ML, Palumbi SR. 2014. Meta-analysis reveals lower genetic diversity in overfished populations. *Mol Ecol*. 23(1):29–39. <https://doi.org/10.1111/mec.12509>.
- Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ. 2009. Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish Fisheries*. 10(4):361–395. <https://doi.org/10.1111/j.1467-2979.2008.00324.x>.
- Riginos C, Liggins L. 2013. Seascape genetics: populations, individuals, and genes marooned and adrift. *Geogr Compass*. 7(3): 197–216. <https://doi.org/10.1111/gec3.12032>.
- Schlenker LS, Stieglitz JD, Greer JB, Faillettaz R, Lam CH, Hoenig RH, Heuer RM, McGuigan CJ, Pasparakis C, Esch EB, et al. 2022. Brief oil exposure reduces fitness in wild Gulf of Mexico mahi-mahi (*Coryphaena hippurus*). *Environ Sci Technol*. 56(18):13019–13028. <https://doi.org/10.1021/acs.est.2c01783>.
- Smit A, Hubley R, Green P. 2013. RepeatMasker. Open-4.0. Available in: <http://www.repeatmasker.org>.
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. 34(Web Server):W435–W439. <https://doi.org/10.1093/nar/gkl200>.
- Sun H, Ding J, Piednoël M, Schneeberger K. 2018. findGSE: estimating genome size variation within human and Arabidopsis using k-mer frequencies. *Bioinformatics*. 34(4):550–557. <https://doi.org/10.1093/BIOINFORMATICS/BTX637>.
- Tan MH, Austin CM, Hammer MP, Lee YP, Croft LJ, Gan HM. 2018. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellularis*) genome assembly. *Gigascience*. 7(3):1–6. <https://doi.org/10.1093/gigascience/gix137>.
- Vij S, Kuhl H, Kuznetsova IS, Komissarov A, Yurchenko AA, Van Heusden P, Singh S, Thevasagayam NM, Prakki SRS, Purushothaman K, et al. 2016. Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *PLoS Genet*. 12(4):e1005954. <https://doi.org/10.1371/JOURNAL.PGEN.1005954>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 9(11):e112963. <https://doi.org/10.1371/journal.pone.0112963>.

Editor: C. Palaikostas