



Check for updates

GENOME NOTE

Chromosome-level genome assembly of the hamour (orange-spotted grouper), *Epinephelus coioides*

[version 1; peer review: 2 approved, 1 approved with reservations]

Razan Khalifa^{1,2}, Tim Bean ³, Dana Albatesh^{1,3}, Ronny van Aerle ⁴,
Zenaba Kahtir¹, Zainab Hizam ¹, Marta Gut^{5,6}, Francisco Câmara^{5,6},
Fernando Cruz ^{5,6}, Jèssica Gómez-Garrido^{5,6}, Tyler Alioto ^{5,6}, Eduarda Santos²,
Alexandra Leitão¹, Diana Minardi ⁴

¹Environmental Science Center, Qatar University, Doha, Qatar²University of Exeter College of Life and Environmental Sciences, Exeter, England, UK³The University of Edinburgh The Roslin Institute, Roslin, Scotland, UK⁴Centre for Environment Fisheries and Aquaculture Science, Weymouth, England, UK⁵Centro Nacional de Análisis Genómico, Barcelona, Spain⁶Universitat de Barcelona, Barcelona, Spain

V1 First published: 29 Oct 2025, 14:1180
<https://doi.org/10.12688/f1000research.153918.1>

Latest published: 29 Oct 2025, 14:1180
<https://doi.org/10.12688/f1000research.153918.1>

Abstract

We present a chromosome-level genome assembly and annotation of the hamour, or orange-spotted grouper (*Epinephelus coioides*), a high-value and significant teleost fish species across West Indo-Pacific regions of the Middle East, South Africa, and Australia. This species is a popular target for both commercial and recreational fishing and it is widely cultured around the world, particularly in the Asia-Pacific region. The hamour genome was sequenced from one individual male originating from a wild population in the Arabian Gulf and assembled into a 1.07 Gb assembly, the largest 24 superscaffolds making up 99.9% of the assembly. Annotation of the genome identified 28,384 protein-coding genes, with 98.9% single-copy BUSCO gene completeness (Actinopterygii database). These data will support further studies on functional ecological and evolutionary genomics of this species, enhancing the understanding of its biology and its responses to stressors including pathogens.

Keywords

Aquaculture, Nanopore, Illumina, Arabian Gulf, teleost, Serranidae, Qatar, whole genome sequencing

Open Peer Review

Approval Status

	1	2	3
version 1			
29 Oct 2025	view	view	view

1. **Jayan Duminda M Senevirathna**, Uva Wellassa University, Badulla, Sri Lanka

2. **Merly Escalona**, University of Connecticut Department of Molecular and Cell Biology, Storrs, USA

3. **Karim Gharbi** , The Earlham Institute, Norwich, UK

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the **Genomics and Genetics** gateway.

Corresponding author: Diana Minardi (diana.minardi@cefas.gov.uk)

Author roles: **Khalifa R:** Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Bean T:** Conceptualization, Methodology, Resources, Supervision, Writing – Review & Editing; **Albatesh D:** Investigation; **van Aerle R:** Data Curation, Resources, Writing – Review & Editing; **Kahtir Z:** Investigation; **Hizam Z:** Investigation; **Gut M:** Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – Original Draft Preparation; **Câmara F:** Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – Original Draft Preparation; **Cruz F:** Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – Original Draft Preparation; **Gómez-Garrido J:** Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – Original Draft Preparation; **Alioto T:** Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – Original Draft Preparation; **Santos E:** Conceptualization, Supervision, Writing – Review & Editing; **Leitão A:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Minardi D:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This publication was made possible by the MME02-0924-200032 award “Sustainable aquaculture Qatar (SAQ): Understanding the threat posed by emerging aquatic animal diseases” from the Qatar National Research Fund (QNRF) a member of Qatar Foundation (QF)/Qatar Research Development and Innovation Council (QRDI). The findings herein reflect the work and are solely the responsibility of the authors. The work was also supported by Roslin Institute Strategic Programme Award BBS/E/RL/230001A “Identifying the genomic basis of complex traits in farmed animals”.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Khalifa R *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Khalifa R, Bean T, Albatesh D *et al.* **Chromosome-level genome assembly of the hamour (orange-spotted grouper), *Epinephelus coioides* [version 1; peer review: 2 approved, 1 approved with reservations]** F1000Research 2025, 14:1180 <https://doi.org/10.12688/f1000research.153918.1>

First published: 29 Oct 2025, 14:1180 <https://doi.org/10.12688/f1000research.153918.1>

Introduction

Epinephelus coioides (Hamilton, 1822), commonly known as the orange-spotted grouper or hamour in the Arabian Gulf region, is a fish species that belongs to the family Serranidae (Rimmer and Glamuzina, 2019). It is widely distributed across the Indo-Pacific region, including the coastal waters of the Arabian Gulf, comprising Qatar's coastal regions. This species is highly valued for its meat quality, is a popular target for both commercial and recreational fishing, and is widely cultured around the world particularly in the Asia-Pacific region (Ranjan et al., 2017). While hatchery production has been increasing to mitigate the pressures on wild populations (Antoro et al., 2006), juveniles are still being captured from the wild for mariculture operations (Tupper and Sheriff, 2008), with potential impacts for wild population sustainability.

The hamour presents a distinctive appearance with an elongated body and a large mouth. Its coloration can vary, but it typically features a brownish or reddish background adorned with prominent orange or reddish spots, giving it its common name. It has a robust build and can reach large sizes, with adult specimens sometimes exceeding one meter in length (Chen et al., 2005). This species constitutes a highly valued commodity, reaching substantial prices in the international market, for example ranging from 11.70 to 40.30 USD per kilogram with an average of 21.45 USD per kilogram in Hong Kong and south east Asia (Ranjan et al., 2017). In the Arabian Gulf, it is the most important reef-associated commercial species, collected via demersal traps, longlines, and trawls (Grandcourt et al., 2009). In northern Oman and Iran, it is targeted using wire basket traps and is also commonly taken as bycatch in trawl fisheries (McIlwain et al., 2016).

Genome sequencing plays a vital role in understanding the genetic makeup of a species, including both genes and regulatory elements, and facilitates understanding of its evolutionary history. By analysing its genome, insights into multiple aspects of a species' biology, such as disease resistance, growth patterns, and reproductive characteristics can be obtained. To date, several genomics studies have been performed in this species, for example, genome-wide association studies were used to explore ammonia tolerance (Xu et al., 2019), a highly flexible and repeatable single nucleotide polymorphism genotyping method was employed to study its growth and ammonia tolerance (Shan et al., 2023), and whole genome sequencing and analysis revealed key regulatory pathways influencing sex differentiation (Li et al., 2023).

Here, we present a highly continuous chromosome-level genome assembly of this species, obtained using long and short read sequencing technologies and Omni-C scaffolding.

Methods

Sampling and nucleic acids extractions

One individual male hamour (43 cm) was caught from the wild (East Qatar, Arabian Gulf, September 2022), and sampled directly. For genomic DNA extraction, 50 mg of liver tissue were excised, submerged in isopentane bath (prepared over dry ice) until frozen (30 s), stored in a pre-chilled cryotube and preserved at -80 °C until further analyses. In parallel, 50 mg of tissue from liver, kidney, heart, spleen, gonad, muscle, skin, gill, and tail were collected from the same specimen and placed in cryotubes containing RNeasy Lysis Buffer (Qiagen), stored at +4 °C for 24 hrs to ensure appropriate tissue preservation, and then moved to -80 °C until RNA extraction.

High molecular weight (HMW) DNA was extracted from the liver sample using QIAGEN® Genomic tip Blood & Cell Culture DNA Midi Kit (QIAGEN) following the manufacturer's protocol. Purified DNA was quantified using an Invitrogen™ Qubit™ DNA BR Assay kit (Thermo Fisher Scientific), the purity of the sample was analysed using a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific) UV/Vis, and the integrity of the DNA was assessed with Femto Pulse Genomic DNA 165 kb kit (Agilent). DNA was stored at +4 °C until library preparation.

Total RNA was extracted using a Promega Maxwell® RSC 48 Instrument and Maxwell RSC simply RNA Tissue kit (Promega), quantified using an Invitrogen Qubit RNA BR Assay kit (Thermo Fisher Scientific), and analysed for purity using an RNA 6000 Nano Bioanalyzer 2100 Assay (Agilent). Purified RNA samples were stored at -80 °C until library preparation.

Library preparation and sequencing

A long-read library for Oxford Nanopore Technologies™ (ONT) sequencing was prepared from DNA extracted from the liver using the 1D Sequencing kit SQK-LSK110 (ONT). In brief, 3 µg of DNA underwent end-repair and adenylation using the NEBNext® Ultra™ II End Repair/dA-Tailing Module (New England Biolabs), followed by ligation of sequencing adaptors. The ligation product was purified using Beckman Coulter™ AMPure XP Beads (Beckman Coulter Life Science) and eluted in Elution Buffer (ONT). The library was sequenced on a PromethION™ 24 instrument with a R9.4.1 flow cell, and data collected for 110 hours. The quality parameters of the sequencing run were monitored in real time using the MinKNOW™ platform v22.10.7 (ONT, <https://nanoporetech.com/document/experiment-companion-minknow>) and basecalling performed using Guppy v6.3.9 (ONT, <https://nanoporetech.com/document/Guppy-protocol>).

For the proximity ligation library, the Dovetail® Omni-C® Kit (Dovetail Genomics, Cantata Bio) was used on the HMW DNA extracted from the liver, following the manufacturer's protocol. After reversal crosslinking, the DNA was purified and biotinylated chimeric molecules isolated using streptavidin beads before PCR enrichment with 12 PCR cycles using KAPA HiFi HotStart Ready Mix (Roche). The short-insert paired-end library for whole genome sequencing was prepared using the PCR-free protocol and the KAPA HyperPrep kit (Roche). After end-repair and adenylation, Illumina™ platform-compatible adaptors with unique dual indexes and unique molecular identifiers (Integrated DNA Technologies) were ligated. The sequencing library was quality controlled on a 2100 Bioanalyzer using the DNA 7500 assay (Agilent), quantified with KAPA Library Quantification Kit (Roche), and sequenced on four lanes of a NovaSeq™ 6000 (Illumina) with a read length of 2×151 bp.

Total RNA from individual tissues (tail, gonad, heart, gill, skin, spleen, liver, kidney, and muscle) was used to prepare RNA-Seq tissue-specific libraries with a KAPA Stranded mRNA-Seq kit (Roche) following the manufacturer's protocol. The transcriptomes were sequenced on 4 lanes of a NovaSeq 6000 (Illumina) with a read length of 2×151 bp.

Nuclear genome assembly, Omni-C scaffolding, curation, and assembly quality checks

Data generated with the PromethION and NovaSeq 6000 were assembled with the Centro Nacional de Análisis Genómico (CNAG) Snakemake pipeline v2.0 (https://github.com/cnag-aat/assembly_pipeline) to obtain an optimal base assembly for further Omni-C scaffolding. The list of programs, parameters and versions used to assemble and quality check the genome are presented in Table 1. In brief, Illumina reads were processed with Cutadapt (Martin, 2011), while ONT reads were filtered with FiltLong (<https://github.com/rrwick/FiltLong>). Filtered ONT reads were assembled with both Flye (Kolmogorov et al., 2019) and NextDenovo (Hu et al., 2024). GenomeScope2 (Ranallo-Benavidez et al., 2020; Vurture et al., 2017) was used to estimate genome size with the 20-mers present in the pre-processed Illumina reads. The NextDenovo (Hu et al., 2024) assembly was polished with both ONT and Illumina paired-end reads using Hypo (Kundu et al., 2019) and then the polished assembly was collapsed with purge_dups (Guan et al., 2020) to remove haplotypic duplications.

Table 1. Programs with citations, versions and parameters used in the present study. Dark grey: genome assembly. Light grey: genome annotation. Dark blue: assembly checks. Light blue: genome curation.

Program	Version	Parameters ^A and notes
Augustus (Stanke et al., 2006)	3.5.0	
BEDtools (Quinlan and Hall, 2010)	2.29.0	
BLAST (Altschul et al., 1990)	2.12.0	Against UniProt (May 2023)
BlobToolKit (Challis et al., 2020)	4.1.5	
BUSCO (Manni et al., 2021)	5.7.1	-m genome (odb_10 Actinopterygii, Fungi, and Bacteria)
Cutadapt* (Martin, 2011)	4.1	-q 20 --paired --retain_unpaired
Dovetail Genomics https://omni-c.readthedocs.io/en/latest/fastq_to_bam.html		-mq 40
ESPRESSO (Gao et al., 2024)	1.3.0	
EVidenceModeler (Haas et al., 2008)	1.1.1	
fasta-stats.py https://github.com/cnag-aat/scripts/blob/main/fasta-stats.py		
FiltLong* https://github.com/rrwick/FiltLong	0.2.1	--min_length 1000 --min_mean_q 80
Flye* (Kolmogorov et al., 2019)	2.9.1-b1780	--nano-raw -i 2 --scaffold -g 2g
GeneID (Alioto et al., 2018)	1.4	
Genemark-ET (Lomsadze et al., 2014)	4.71	
GenomeScope2* (Ranallo-Benavidez et al., 2020; Vurture et al., 2017)	2	
Hypo* (Kundu et al., 2019)	1.0.3	-c 100.62178934949716 -s 600m
mercury* (Rhie et al., 2020)	1.3	k=19
minimap2* (Li, 2018)	2.24-r1122	-ax map-ont

Table 1. *Continued*

Program	Version	Parameters [^] and notes
miniprot (Li, 2023)	0.6	
NextDenovo* (Hu et al., 2024)	2.5.0	read_cutoff=1k genome_size=600m seed_depth=45 seed_cutoff=0 blocksize=1g
PANNZER (Törönen and Holm, 2022) http://ekhidna2.biocenter.helsinki.fi/sanspanz/		
PASA (Haas et al., 2008)	2.5.2	
PretextGraph https://github.com/sanger-tol/PretextGraph	0.0.6	
PretextView https://github.com/sanger-tol/PretextView	0.2.5	
purge_dups* (Guan et al., 2020)	1.2.5	cutoffs -l 5 -m 192 -u 576
RepeatMasker (Smith et al., 2007) http://www.repeatmasker.org	4.1.2	
RepeatModeler https://github.com/Dfam-consortium/RepeatModeler	1.0.11	
SAMtools (Danecek et al., 2021; Li et al., 2009)	1.9	
STAR (Dobin et al., 2013)	2.7.2a	
StringTie (Pertea et al., 2015)	2.2.1	
TACO (Niknafs et al., 2017)	0.7.3	
telomeric-identifier (Brown et al., 2023)	0.2.41	
TransDecoder https://github.com/TransDecoder/TransDecoder	5.7.1	
YaHS* (Zhou et al., 2023)	1.2a.2	

*Program ran within the Centro Nacional de Análisis Genómico (CNAG) snakemake pipeline v2.0 (https://github.com/cnag-aat/assembly_pipeline).

[^]If different from default parameters.

For further proximity ligation-based scaffolding, a total of 206.96 million Omni-C read pairs were mapped to the assembled genome using the Dovetail Genomics recommended protocol (https://omni-c.readthedocs.io/en/latest/fastq_to_bam.html). After excluding PCR duplicates, 106.91 million valid Omni-C read pairs were used to scaffold the assembly with YaHS (Zhou et al., 2023) using the default initial contig error correction step.

To guide manual curation of the assembly, the ONT read coverage was computed for all positions in the assembly using minimap2 (Li, 2018), SAMtools (Danecek et al., 2021; Li et al., 2009), and BEDtools (Quinlan and Hall, 2010), as well as the location of gaps with fasta-stats.py (<https://github.com/cnag-aat/scripts/blob/main/fast-stats.py>) and telomeres with telomeric-identifier (Brown et al., 2023). These extensions were added to the contact map using PretextView (<https://github.com/sanger-tol/PretextView>). Manual curation was performed according to the rapid curation protocol from The Sanger Institute (<https://gitlab.com/wtsi-grit/rapid-curation>) using PretextView (<https://github.com/sanger-tol/PretextView>). The genome was assessed for completeness with BUSCO using the odb10 Actinopterygii database (Manni et al., 2021), with Merquy (Rhie et al., 2020) for consensus accuracy (QV) and k-mer statistics, for contiguity statistics with fasta-stats.py (<https://github.com/cnag-aat/scripts/blob/main/fast-stats.py>), and for contamination with Blob-ToolKit (with NCBI nt database, August 2023 update) (Challis et al., 2020) and BUSCO using the odb10 databases for Fungi and Bacteria (Manni et al., 2021). For comparison with the genome assembled in this study, the genome previously described by Li and colleagues (Li et al., 2023) and available in the European Nucleotide Archive (accession ID: PRJEB28248) was also assessed for completeness with BUSCO's odb10 Actinopterygii database (Manni et al., 2021).

Genome annotation

The hamour genome assembly annotation was obtained by combining transcript alignments, protein alignments and *ab initio* gene predictions. The list of programs, parameters, and versions used for genome annotation is provided in Table 1. In brief, repeats present in the genome assembly were annotated with RepeatMasker (Smith et al., 2007; <http://www.repeatmasker.org>) using the custom repeat library available for *Danio rerio* and a new repeat library specific for this study made with RepeatModeler (<https://github.com/Dfam-consortium/RepeatModeler>). After excluding repeats

that were part of repetitive protein families from the resulting library, RepeatMasker (Smith et al., 2007; <http://www.repeatmasker.org>) was run again with this new library performing a BLAST (Altschul et al., 1990) search against UniProt (May 2023, <https://www.uniprot.org/>) to annotate the specific repeats. RNA-seq reads were aligned to the previously assembled genome using STAR (Dobin et al., 2013). Transcript models were subsequently generated using StringTie (Pertea et al., 2015) and merged using TACO (Niknafs et al., 2017). High-quality junctions to be used during the annotation process were obtained by running ESPRESSO (Gao et al., 2024) after mapping with STAR (Dobin et al., 2013). Finally, assembled spliced alignments were produced with PASA (Haas et al., 2008). TransDecoder (<https://github.com/TransDecoder/TransDecoder>) was run on the spliced alignments in PASA (Haas et al., 2008) to detect coding regions in the transcripts. The complete proteomes of *Gymnodraco acuticeps*, *Sander lucioperca*, *Cottoperca gobio*, and *Perca fluviatilis* were downloaded from UniProt (May 2023, <https://www.uniprot.org/>) and aligned to the genome using miniprot (Li, 2023). *Ab initio* gene predictions were performed on the repeat-masked assembly with GeneID (Alioto et al., 2018) and Augustus (with human parameters) (Stanke et al., 2006), and Genemark-ET in self-trained mode (Lomsadze et al., 2014) with and without incorporating evidence from the RNA-seq data. Finally, all the data were combined into consensus coding sequence models using EVidenceModeler (Haas et al., 2008). Additionally, untranslated regions (UTRs) and alternative splicing forms were annotated via two rounds of PASA (Haas et al., 2008) annotation updates. Functional annotation was performed on the annotated proteins with PANNZER's online server (Törönen and Holm, 2022; <http://ekhidna2.biocenter.helsinki.fi/sanspanz/>).

Results, Discussion, and Conclusions

ONT whole genome sequencing produced 137.75 Gb of data (coverage=128.62x) and Illumina produced 67.73 Gb of 2x151 bp pair-end reads (coverage=63.24x). Genome size (genome haploid length) estimated with GenomeScope2 ranged from 1,088,845,762 to 1,089,817,901 bp (Table 2). For proximity ligation-based scaffolding, a total of 206.96 million Omni-C read pairs were mapped to the intermediate assemblies generated with NextDenovo (Hu et al., 2024), resulting in a final assembly with scaffold N50 of 45.64 Mb, N90 of 39.86 Mb and accounting for 1.07 Gb (Table 3,

Table 2. Hamour genome assembly size. Genome size estimated by GenomeScope2 (Ranallo-Benavidez et al., 2020; Vurture et al., 2017) on the pre-processed Illumina reads. bp: base pairs.

Attribute	Minimum	Maximum
Homozygous	99.64%	99.66%
Heterozygous	0.34%	0.36%
Genome Haploid Length (bp)	1,088,845,762	1,089,817,901
Genome Repeat Length (bp)	320,257,215	320,543,145
Genome Unique Length (bp)	768,588,547	769,274,755
Model Fit	74.61%	98.69%
Read Error Rate	0.14%	0.14%

Table 3. Hamour genome assembly results and comparison with current publicly available genome (Li et al., 2023). Genome completeness was assessed with BUSCO (Manni et al., 2021) using the Actinopterygii odb10 database updated on the 08/01/2024. Number of BUSCO groups searched for in the Actinopterygii database was 3,640. bp: base pairs; Mb: megabases.

Attribute	This study	Li et al., 2023
Genome assembly total length (bp)	1,071,864,792	1,023,559,032
Scaffolds number	33	1450
Scaffold N50 (Mb)	45	2
Contigs number	140	159
Contig N50 (Mb)	18	2
Completeness	99.3%	99.0%
Single-copy	98.9%	98.5%
Duplicated	0.4%	0.5%
Fragmented	0.6%	0.7%
Missing	0.1%	0.3%

Figure 1), consistent with the GenomeScope2 (Ranallo-Benavidez et al., 2020; Vurture et al., 2017) estimation. The assembled genome consists of 24 superscaffolds (making up 99.9% of the assembly) in accordance with the previously reported diploid karyotype ($2n=48$) for this species (Wang et al., 2010). It had a consensus accuracy of QV=47 and single-copy BUSCO gene completeness of 98.9% (BUSCO odb10 Actinopterygii) (Table 3). No evidence of contamination was detected. Together, these statistics indicate that we have assembled a high quality, chromosome-level genome for the hamour.

A comparison of the chromosome-level genome assembly produced in this study and a previously published genome (Li et al., 2023) is presented in Table 3. Genome size was consistent in both studies (1.07 and 1.02 GB in the present and previous study, respectively). Our assembly had a lower number of scaffolds with a greater scaffold length, providing an improvement on the previously published genome and contributing to the advancement of research for this species.

The genome annotation identified 28,384 protein-coding genes, producing 39,296 transcripts (1.38 transcripts per gene), which improved on the number of annotated protein coding genes reported in Li et al. (2023) (26,931). The annotated transcripts contained 11.05 exons on average, with 91% of them being multi-exonic (Table 4).

Scaffold statistics

- Log10 scaffold count (total 33)
- Scaffold length (total 1.1G)
- Longest scaffold (53M)
- N50 length (46M)
- N90 length (40M)

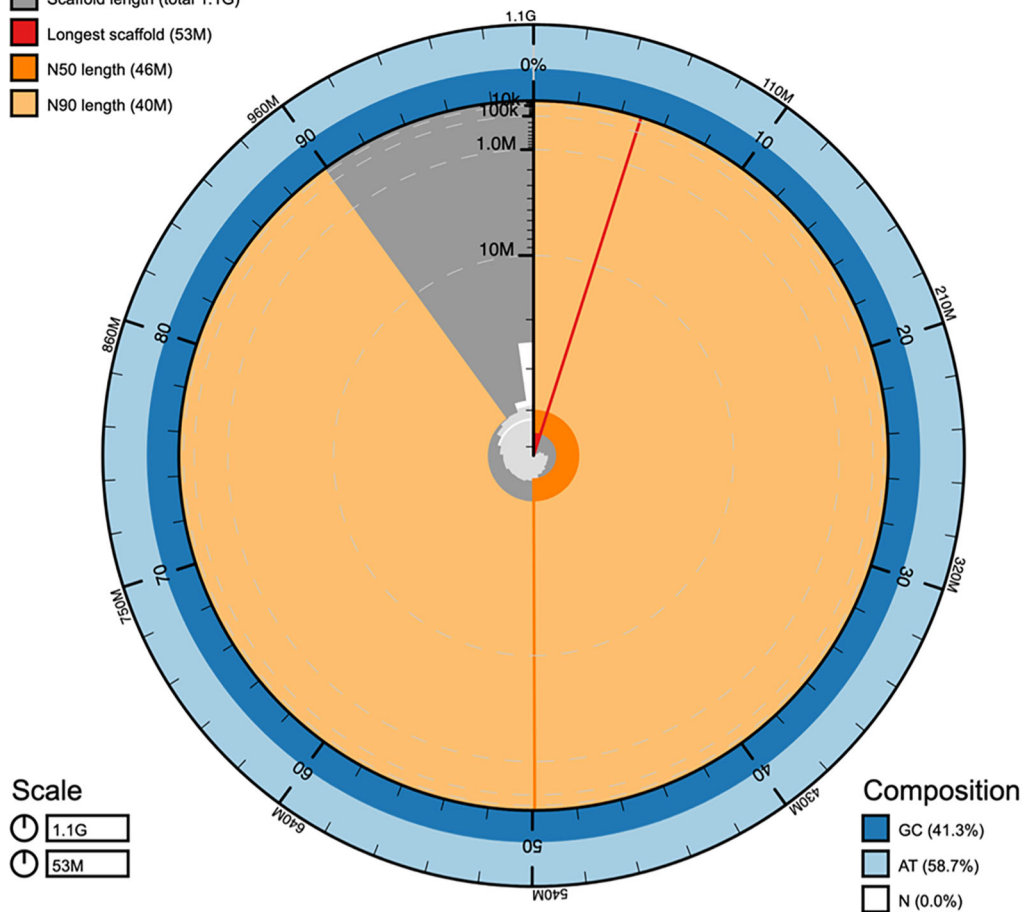


Figure 1. Snail plot summary of assembly statistics for the hamour genome assembly produced in this study. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,071,864,792 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (52,562,209 bp, in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (45,643,039 and 39,861,643 bp respectively). The light grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and light-blue area around the outside of the plot shows the distribution of bases GC, AT, and N % in the same bins as the inner plot. bp: base pair; GC: guanine-cytosine; AT: adenine-thymine; N: nucleobase.

Table 4. Genome annotation results. The hamour genome assembly annotation was achieved by combining transcript alignments, protein alignments and *ab initio* gene predictions. The genome was annotated with 28,384 protein-coding genes, producing 39,296 transcripts (1.38 transcripts per gene). bp: base pairs; Mb: megabases.

Attribute	Result
Number of protein-coding genes	28,384
Median gene length (bp)	8,711
Number of transcripts	39,296
Number of exons	284,015
Number of coding exons	269,906
Median UTR length (bp)	752
Median intron length (bp)	500
Exons/transcript	11.052
Transcripts/gene	1.38
Multi-exonic transcripts	91%
Gene density (gene/Mb)	26.48

Table 5. Genome data for *Epinephelus coioides* (orange-spotted grouper, hamour).

Project accession data	
Assembly identifier	QU_Ecoi
Species	<i>Epinephelus coioides</i>
Specimen	QU-Ecoi-1
NCBI Taxonomy ID	94232
BioProject	PRJNA1128520
BioSample ID	SAMN42050860, SAMN43492902-SAMN43492913
Isolate information	QU-Ecoi-1
Raw data accessions	
Oxford Nanopore PromethION	SRR30574011
Omni-C Illumina	SRR30574012
Illumina short-read	SRR30574003
Illumina RNASeq	SRR30574004-SRR30574010; SRR30574013-SRR30574014
Genome assembly	
Assembly accession	GCA_051314025.1

Here we report on the sequencing and assembly of a hamour individual from the Arabian Gulf using a combination of Nanopore and Illumina sequencing technologies. We produced a chromosome-level assembly for this species and have improved on its annotation compared to a previously released genome. The genome sequence, raw data, and annotation are released openly for reuse. All raw sequence data, the assembly, and annotations have been deposited in INSDC databases, with accession identifiers reported in [Table 5](#). These data will facilitate further studies on the biology of this species and on its management in the wild and aquaculture settings.

Ethical considerations

Due the nature of the research project, with no experimental work on live animals (working only with tissues collected from dead animals), an exemption certificate from our institutional animal care and use committee (IACUC) was obtained for the use of wild fish caught by independent fishermen and bought by the author immediately after capture. The hamour used in this study was euthanised via immersion in MS-222 and destruction of the brain immediately after being caught. This study followed all relevant ethical guidelines and protocols approved by the institutional animal care and use committee (IACUC).

Data availability

Underlying data

The assembled genome and raw data are publicly available through the National Centre for Biotechnology Information (NCBI) and Short Read Archive (SRA) under the following identifiers: Organism: *Epinephelus coioides* QU-Ecoi-1, Assembly accession: GCA_051314025.1, BioProject: PRJNA1128520, and BioSample: SAMN42050860 (Table 5).

Acknowledgment

Institutional support to CNAG was from the Spanish Government, Ministry of Science, Innovation and Universities and Generalitat de Catalunya through the Departament de Recerca i Universitats and Departament de Salut.

References

- Alioto T, Blanco E, Parra G, *et al.*: **Using geneid to identify genes.** *Curr. Protoc. Bioinformatics.* 2018; **64**: e56.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Altschul SF, Gish W, Miller W, *et al.*: **Basic local alignment search tool.** *J. Mol. Biol.* 1990; **215**: 403–410.
[Publisher Full Text](#)
- Antoro S, Na-Nakorn U, Koedprang W: **Study of genetic diversity of orange-spotted grouper, *Epinephelus coioides*, from Thailand and Indonesia using microsatellite markers.** *Mar. Biotechnol.* 2006; **8**: 17–26.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Brown M, De la Rosa G, Manuel P, *et al.*: **A telomere identification toolkit.** *Zenodo.* 2023.
[Publisher Full Text](#)
- Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – Interactive quality assessment of genome assemblies.** *G3: Genes, Genomes, Genetics.* 2020; **10**: 1361–1374.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chen R, Li W, Lin H: **cDNA cloning and mRNA expression of neuropeptide Y in orange spotted grouper, *Epinephelus coioides*.** *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* 2005; **142**: 79–89.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *Gigascience.* 2021; **10**: giab008.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dobin A, Davis CA, Schlesinger F, *et al.*: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics.* 2013; **29**: 15–21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gao Y, Wang F, Wang R, *et al.*: **ESPRESSO: Robust discovery and quantification of transcript isoforms from error-prone long-read RNA-seq data.** *Sci. Adv.* 2024; **9**: eabq5072.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grandcourt EM, Al Abdessalaam TZ, Francis F, *et al.*: **Reproductive biology and implications for management of the orange-spotted grouper *Epinephelus coioides* in the southern Arabian Gulf.** *J. Fish Biol.* 2009; **74**: 820–841.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**: 2896–2898.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Haas BJ, Salzberg SL, Zhu W, *et al.*: **Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments.** *Genome Biol.* 2008; **9**: R7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hamilton F: *An account of the fishes found in the river Ganges and its branches.* Edinburgh; and Hurst, Robinson, and Co. 90, Cheapside, London: Printed for Archibald Constable and Company; 1822.
[Publisher Full Text](#)
- Hu J, Wang Z, Sun Z, *et al.*: **NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads.** *Genome Biol.* 2024; **25**: 107.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kolmogorov M, Yuan J, Lin Y, *et al.*: **Assembly of long, error-prone reads using repeat graphs.** *Nat. Biotechnol.* 2019; **37**: 540–546.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kundu R, Casey J, Sung W-K: **HyPo: super fast & accurate polisher for long read genome assemblies.** *bioRxiv.* 2019; 2019.12.19.882506.
[Publisher Full Text](#)
- Li H: **Protein-to-genome alignment with miniprot.** *Bioinformatics.* 2023; **39**: btad014.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**: 3094–3100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li H, Handsaker B, Wysoker A, *et al.*: **The sequence alignment/map format and SAMtools.** *Bioinformatics.* 2009; **25**: 2078–2079.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li S, Li W, Jiang S, *et al.*: **Mechanisms of sex differentiation and sex reversal in hermaphrodite fish as revealed by the *Epinephelus coioides* genome.** *Mol. Ecol. Resour.* 2023; **23**: 920–932.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lomsadze A, Burns PD, Borodovsky M: **Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm.** *Nucleic Acids Res.* 2014; **42**: e119–e119.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Manni M, Berkeley MR, Seppely M, *et al.*: **BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol. Biol. Evol.* 2021; **38**: 4647–4654.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet Journal.* 2011; **17**: 10–12.
[Publisher Full Text](#)
- McIlwain J, Ambu-Ali A, Al Jardani N, *et al.*: **Demographic profile of an overexploited serranid, the orange-spotted grouper (*Epinephelus coioides*), from Northern Oman.** *Fish. Bull.* 2016; **114**: 490–502.
[Publisher Full Text](#)
- Niknafs YS, Pandian B, Iyer HK, *et al.*: **TACO produces robust multisample transcriptome assemblies from RNA-seq.** *Nat. Methods.* 2017; **14**: 68–70.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Perlea M, Perlea GM, Antonescu CM, *et al.*: **StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.** *Nat. Biotechnol.* 2015; **33**: 290–295.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**: 841–842.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat. Commun.* 2020; **11**: 1432.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ranjan R, Muktha M, Ghosh S, *et al.*: **Prioritized species for mariculture in India.** Kochi: ICAR - Central Marine Fisheries Research Institute; 2017.
- Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**: 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rimmer MA, Glamuzina B: **A review of grouper (Family Serranidae: Subfamily Epinephelinae) aquaculture from a sustainability science perspective.** *Rev. Aquac.* 2019; **11**: 58–87.
[Publisher Full Text](#)
- Shan X, Zhang X, Ruan Z, *et al.*: **Genomic selection of orange-spotted grouper (*Epinephelus coioides*) based on multiplex PCR enrichment capture sequencing.** *Aquac. Fish.* 2023; **8**: 681–688.
[Publisher Full Text](#)

Smith CD, Edgar RC, Yandell MD, *et al.*: **Improved repeat identification and masking in Dipterans.** *Gene*. 2007; **389**: 1–9.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Stanke M, Schöffmann O, Morgenstern B, *et al.*: **Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources.** *BMC Bioinformatics*. 2006; **7**: 62.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Törönen P, Holm L: **PANNZER—A practical tool for protein function prediction.** *Protein Sci*. 2022; **31**: 118–128.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Tupper M, Sheriff N: **Capture-based aquaculture of groupers.**

Lovatelli A, Holthus PF, editors. *Capture-Based Aquaculture. Global Overview.* *FAO Fisheries Technical Paper*. No. 508. Rome, Italy: 2008; pp. 217–253.

Vurtture GW, Sedlazeck FJ, Nattestad M, *et al.*: **GenomeScope: fast reference-free genome profiling from short reads.** *Bioinformatics*. 2017;

33: 2202–2204.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wang S, Su Y, Ding S, *et al.*: **Cytogenetic analysis of orange-spotted grouper, *Epinephelus coioides*, using chromosome banding and fluorescence in situ hybridization.** *Hydrobiologia*. 2010; **638**: 1–10.

[Publisher Full Text](#)

Xu T, Zhang X, Ruan Z, *et al.*: **Genome resequencing of the orange-spotted grouper (*Epinephelus coioides*) for a genome-wide association study on ammonia tolerance.** *Aquaculture*. 2019; **512**: 734332.

[Publisher Full Text](#)

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics*. 2023; **39**: btac808.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:   

Version 1

Reviewer Report 04 December 2025

<https://doi.org/10.5256/f1000research.168876.r428838>

© 2025 Gharbi K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Karim Gharbi 

The Earlham Institute, Norwich, UK

This genome note reports on the assembly and annotation a chromosome-level reference assembly for the orange-spotted grouper, a popular serranid fish distributed across the Indo-Pacific, and locally known as hamour in the Arabian Peninsula and Persian Gulf region. The note includes a high-level introduction of the species followed by assembly/annotation methods and statistics, with comparison to a previously published assembly from the same species.

The introduction highlights the importance of the species for both commercial and recreational fishers, with a growing aquaculture sector. The sequencing, assembly, and annotation methods are comprehensive and generally follow established practices. The assembly report shows that the authors have achieved chromosome-level scaffolds and significantly improved on the quality of the previous assembly. They have also improved on the number of identified protein-coding genes relative to the previous annotation.

This study represents a valuable contribution to the field of fish genomics, and more specifically species targeted by aquaculture. The availability of a chromosome-level assembly for the hammer will likely have a significant impact on our understanding of the species biology, including conversation and sustainable exploitation.

Below are specific comments, which the authors may want to consider to clarify and improve the manuscript.

Main comments

1. Please comment on the recovery and assembly of the mitochondrial genome.
2. Please clarify that the starting material for Dovetail Omni-C library preparation (currently listed as HMW DNA).
3. Please specify which base-calling model was used in Guppy v6.3.9.
4. Please indicate the type and number of changes introduced during manual curation.

Minor comments

1. Please indicate whether the specimen has been vouchered and include a digital image.
2. Please clarify that the Bioanalyzer assay was used to check the integrity of the RNA extraction.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 25 November 2025

<https://doi.org/10.5256/f1000research.168876.r430760>

© 2025 Escalona M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Merly Escalona

University of Connecticut Department of Molecular and Cell Biology, Storrs, Connecticut, USA

The authors present a set of genomic resources for the orange-spotted grouper. These resources include a chromosome-level genome assembly which metrics meet the benchmark of a high-quality genome, and its corresponding gene annotation.

The genome assembly generated is an improvement on existing genomic resources for the grouper, and extremely welcomed given the evident application of the resource.

The manuscript is overall it is well written and organized as a the type of manuscript submitted, methods are well described and on par with the current pipelines of genome assemblies for the sequencing technologies used in the study.

Having said that, there are a few points in which this manuscript could be improved.

In the introduction, the last two paragraphs are a somewhat disconnected. These should be re-organized and combined, or expanded.

Given the sequencing data generated, it was expected that the nuclear genome assembly was accompanied by a mitochondrial genome as well. Please think about adding this too to the genomic resources.

Information about sequencing output, in particular for long-read data, should be presented with a read distribution and a read length vs. quality plot. These plots are more informative than the single amount of output.

There's no explanation on how k-mer counts were generated for GenomeScope2 and QV metrics.

The assembled genome consists of many scaffolds, but the largest 24 (considered superscaffolds) make up 99.9% of the assembly. The phrase as it is written in the manuscript can be misleading.

You are considering this a chromosome-level assembly, and it has been made available on the INSDC Databases as such. The chromosome naming is not based on ordered sequence length, what's the convention used for the chromosome naming/assignment? was there a specific genome used as reference for the chromosome naming? if so, which one? Please be explicit.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genome assembly, comparative genomics, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 19 November 2025

<https://doi.org/10.5256/f1000research.168876.r430757>

© 2025 Senevirathna J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jayan Duminda M Senevirathna

Uva Wellassa University, Badulla, Uva Province, Sri Lanka

The article is well written, and the content is clearly explained. The availability of raw data adds to the transparency of the study, and the originality of the work is commendable. However, the following comments are for further improvement.

Typically, six or fewer keywords are recommended. Please verify and adjust the number of keywords according to the journal's guidelines.

The Introduction could be strengthened by including additional background information about the fish species, such as sexual dimorphism, ecological significance, and other relevant biological characteristics.

In the Methodology section, please specify that the wild fish samples were obtained from a fisherman. This detail will improve the clarity and reproducibility of your sampling procedure. If possible, include the Mausam voucher number.

If this study presents a chromosome-level assembly, please include a table summarizing the identified chromosomes, along with relevant details such as chromosome length, number of scaffolds anchored, N50 values, and any other assembly statistics that support the quality of the chromosomal organization. Otherwise, change the topic.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Aquatic Biotechnology and Genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research