DATA NOTE

REVISED The genome sequence of the tub gurnard, *Chelidonichthys lucerna* (Linnaeus, 1758) (Perciformes: Triglidae)

[version 2; peer review: 2 approved]
Previously titled : "The genome sequence of the red gurnard, *Chelidonichthys lucerna* (Linnaeus, 1758) (Perciformes: Triglidae)"

Rachel Brittain[1], Patrick Adkins[1], Kesella Scott-Somme[1], Joanna Harley[1], Vengamanaidu Modepali[1],
Marine Biological Association Genome Acquisition Lab,
Darwin Tree of Life Barcoding Collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]The Marine Biological Association, Plymouth, England, UK

v2 **First published:** 06 Jan 2026, **11**:12
https://doi.org/10.12688/wellcomeopenres.25417.1
Latest published: 18 Mar 2026, **11**:12
https://doi.org/10.12688/wellcomeopenres.25417.2

Abstract

We present a genome assembly from an individual *Chelidonichthys lucerna* (tub gurnard; Chordata; Actinopteri; Perciformes; Triglidae). The assembly contains two haplotypes with total lengths of 649.07 megabases and 651.58 megabases. Most of haplotype 1 (96.66%) is scaffolded into 24 chromosomal pseudomolecules. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 16.52 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Chelidonichthys lucerna; red gurnard; genome sequence; chromosomal; Perciformes

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 18 Mar 2026	✓ view	
version 1 06 Jan 2026	✓ view	✓ view

1. **Bingpeng Xing**, Third Institute of Oceanography Ministry of Natural Resources, Xiamen, China

2. **Ivan Rodrigo Wolf**, UNC Charlotte Department of Biological Sciences, Charlotte,

This article is included in the Tree of Life gateway.

USA

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

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Brittain R: Investigation, Resources; **Adkins P**: Writing – Original Draft Preparation, Writing – Review & Editing; **Scott-Somme K**: Investigation, Resources; **Harley J**: Investigation, Resources; **Modepali V**: Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Brittain R, Adkins P, Scott-Somme K *et al.* **The genome sequence of the tub gurnard, *Chelidonichthys lucerna* (Linnaeus, 1758) (Perciformes: Triglidae) [version 2; peer review: 2 approved]** Wellcome Open Research 2026, **11**:12 https://doi.org/10.12688/wellcomeopenres.25417.2

First published: 06 Jan 2026, **11**:12 https://doi.org/10.12688/wellcomeopenres.25417.1

REVISED **Amendments from Version 1**

In Version 2, we have made corrections in response to peer reviewer comments. These include correction of the species binomial and use of the abbreviated species name, and clarification of the interpretation of Merqury.FK k-mer completeness values for each haplotype. The introduction text incorrectly referred to the red gunard, and we are supplying replacement text and references.

Any further responses from the reviewers can be found at the end of the article

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Actinopterygii; Actinopteri; Neopterygii; Teleostei; Osteoglossocephalai; Clupeocephala; Euteleosteomorpha; Neoteleostei; Eurypterygia; Ctenosquamata; Acanthomorphata; Euacanthomorphacea; Percomorphaceae; Eupercaria; Perciformes; Triglioidei; Triglidae; *Chelidonichthys*; *Chelidonichthys lucerna* (Linnaeus, 1758) (NCBI:txid2576622).

Background

The tub gurnard *Chelidonichthys lucerne,* also known as yellow gurnard, is a demersal species found in various benthic substrates such as sandy, muddy, and gravel bottoms, at depths ranging from 20 m to 300 m (Froese & Pauly, 2017; El-Serafy *et al*., 2015; Riedl, 1991). It often co-occurs with other Triglidae with commercial importance, such as the grey gurnard (*Eutrigla gurnardus*), in the Atlantic and Mediterranean (Froese & Pauly, 2017; Lopez-Lopez *et al*., 2011; Colloca *et al*., 2019; ICES, 2007).

The tub gurnard is the largest species of the European Triglidae, reaching a maximum size of 75 cm and living up to 15 years (Lopez-Lopez *et al.,* 2011; Baron, 1985). It exhibits seasonal migratory movements, with a higher concentration in shallower depths during spring and summer and moving to deeper waters in the winter (Campos et al., 2022). In northern regions, tub gurnards spawn between May and July, and males reach sexual maturity at a smaller size and younger age than females (Colloca *et al*., 2003; Vallisneri *et al.,* 2011; Papaconstantinou, 1984; Baron, 1985).

The tub gurnard has high nutritional value (Froese & Pauly, 2017), and has emerged as a species of commercial importance in Europe (ICES, 2007; Deval *et al.,* 2007). In 2006, ICES categorised *C. lucerne* as a species that might be vulnerable to commercial exploitation, hence recommending monitoring programmes to gather biological parameters for stock assessment purposes (ICES, 2007). There is little published information on tub gurnard stocks in UK coastal waters other than distribution data, size frequencies, and length-weight relations (Coull et al., 1989; Parker-Humphreys, 2004, 2005).

We present a chromosome-level genome sequence for *C. lucerna*, one of only two genomes for the genus *Chelidonichthys* as of February 2026 (data obtained via NCBI datasets, O'Leary *et al.*, 2024). Only three genomes are available for the family Triglidae. The assembly was produced using the Tree of Life pipeline from a specimen collected off the coast of Cornwall, UK (Figure 1). This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Darwin Tree of Life Project Consortium, 2022).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult *C. lucerna* (specimen ID MBA-220324-005A, ToLID fCheLuc1; Figure 1), collected by trawl from the *RV Sepia* in Whitsand Bay, Cornwall, UK (latitude 50.3264, longitude −4.2366) on 2022-03-24. The specimen was collected by a group from the Marine Biological Association and formally identified by Rachel Brittain. The same specimen was used for RNA sequencing. For the Darwin Tree of Life sampling and metadata approach, refer to Lawniczak *et al.* (2022).

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode

Figure 1. Photograph of the *Chelidonichthys lucerna* (fCheLuc1) specimen used for genome sequencing.

region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The fCheLuc1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the gill was homogenised by powermashing using a PowerMasher II tissue disruptor.

HMW DNA was extracted in the WSI Scientific Operations core using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by manual SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 32.04 ng/μL and a yield of 1 441.80 ng, with a fragment size of 10.1 kb. The 260/280 spectrophotometric ratio was 1.89, and the 260/230 ratio was 2.06.

RNA was extracted from gill tissue of fCheLuc1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol. The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25 M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking.

The Hi-C sample was prepared from 20–50 mg of frozen gill tissue of the fCheLuc1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing.

Biotinylated DNA constructs were fragmented using a Covaris E220 sonicator and size selected to 400–600 bp using SPRISelect beads. DNA was enriched with Arima-HiC v2 kit Enrichment beads. End repair, A-tailing, and adapter ligation were carried out with the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs), following a modified protocol where library preparation occurs while DNA remains bound to the Enrichment beads. Library amplification was performed using KAPA HiFi HotStart mix and a custom Unique Dual Index (UDI) barcode set (Integrated DNA Technologies). Depending on sample concentration and biotinylation percentage determined at the crosslinking stage, libraries were amplified with 10–16 PCR cycles. Post-PCR clean-up was performed with SPRISelect beads. Libraries were quantified using the AccuClear Ultra High Sensitivity dsDNA Standards Assay Kit (Biotium) and a FLUOstar Omega plate reader (BMG Labtech).

Prior to sequencing, libraries were normalised to 10 ng/µL. Normalised libraries were quantified again to create equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq X.

RNA library preparation and sequencing

Libraries were prepared using the NEBNext ® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo (dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq X to generate 150-bp paired-end reads.

Genome assembly

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts ($k = 31$) was generated from the filtered reads using FastK. GenomeScope2 (Ranallo-Benavidez *et al.*, 2020) was used to analyse the *k*-mer frequency distributions, providing estimates of genome size, heterozygosity, and repeat content.

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021; Cheng *et al.*, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data in YaHS (Zhou *et al.*, 2023), using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 9 breaks and 21 joins. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020) was run in a Singularity container (Kurtzer *et al.*, 2017) to evaluate *k*-mer completeness and assembly quality for both haplotypes using the *k*-mer database ($k = 31$) computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *C. lucerna* specimen generated 78.89 Gb (gigabases) from 7.50 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 593.08 Mb, with a heterozygosity of 0.83% and repeat content of 17.37% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 127× coverage. Hi-C sequencing produced 329.24 Gb from 2 180.43 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 649.07 Mb in 523 scaffolds, with 312 gaps, and a scaffold N50 of 27.99 Mb (Table 2).

Most of the assembly sequence (96.66%) was assigned to 24 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

The mitochondrial genome was also assembled. This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

For haplotype 1, the estimated QV is 61.4, and for haplotype 2, 62.1. When the two haplotypes are combined, the assembly achieves an estimated QV of 61.7. The *k*-mer completeness is 84.28% for haplotype 1, 84.25% for haplotype 2, and 99.62% for the combined haplotypes (Figure 4).

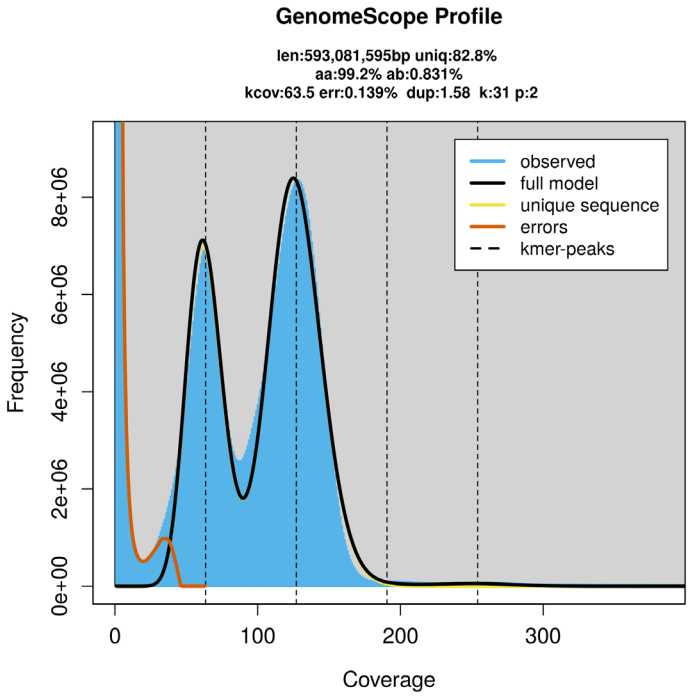

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB74613.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	fCheLuc1	fCheLuc1	fCheLuc1
Specimen ID	MBA-220324-005A	MBA-220324-005A	MBA-220324-005A
BioSample (source individual)	SAMEA112765707	SAMEA112765707	SAMEA112765707
BioSample (tissue)	SAMEA112765787	SAMEA112765786	SAMEA112765786
Tissue	gill	gill	gill
Instrument	Revio	Illumina NovaSeq X	Illumina NovaSeq X
Run accessions	ERR12875161	ERR12893002; ERR12893003	ERR12893004
Read count total	7.50 million	2 180.43 million	55.73 million
Base count total	78.89 Gb	329.24 Gb	8.41 Gb

BUSCO analysis using the actinopterygii_odb10 reference set (n = 3 640) identified 99.3% of the expected gene set (single = 98.3%, duplicated = 1.0%) for haplotype 1. The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the haplotype 1, is **6.C.Q61**, meeting the recommended reference standard.

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**,

Table 2. Genome assembly statistics.

Assembly name	fCheLuc1.hap1.1	fCheLuc1.hap2.1
Assembly accession	GCA_965644535.1	GCA_965644425.1
Assembly level	chromosome	scaffold
Span (Mb)	649.07	651.58
Number of chromosomes	24	scaffold-level
Number of contigs	835	650
Contig N50	3.03 Mb	3.39 Mb
Number of scaffolds	523	345
Scaffold N50	27.99 Mb	27.9 Mb
Longest scaffold length (Mb)	36.43	-
Organelles	Mitochondrion: 16.52 kb	-

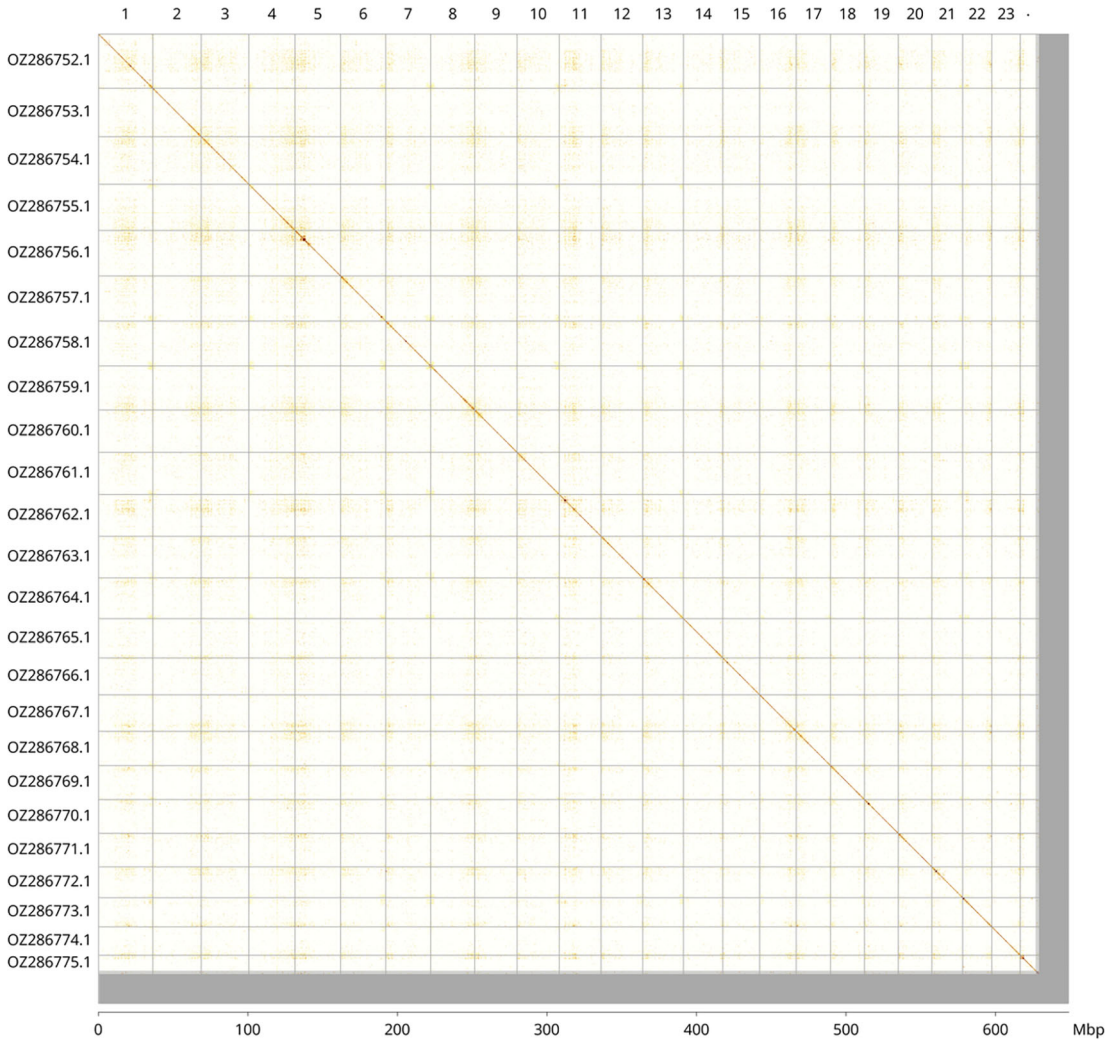

Figure 3. Hi-C contact map of the *C. lucerna* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *C. lucerna* fCheLuc1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ286752.1	1	36.43	42.50
OZ286753.1	2	32.41	42.50
OZ286754.1	3	31.88	42.50
OZ286755.1	4	30.93	42.50
OZ286756.1	5	30.46	43
OZ286757.1	6	30.17	42.50
OZ286758.1	7	29.98	42.50
OZ286759.1	8	29.61	43
OZ286760.1	9	28.28	43
OZ286761.1	10	28.19	43
OZ286762.1	11	27.99	43.50
OZ286763.1	12	27.78	42.50
OZ286764.1	13	27.34	42.50
OZ286765.1	14	26.23	42.50
OZ286766.1	15	24.64	42.50
OZ286767.1	16	24.51	43
OZ286768.1	17	22.80	43
OZ286769.1	18	22.76	43
OZ286770.1	19	22.71	42.50
OZ286771.1	20	22.36	43.50
OZ286772.1	21	20.57	43
OZ286773.1	22	19.49	43
OZ286774.1	23	19.05	43.50
OZ286775.1	24	10.79	45.50

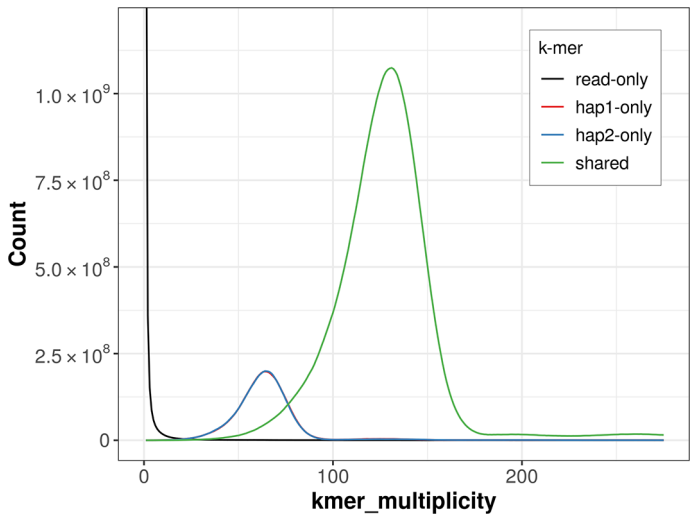

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers that appear in the reads but are not assembled. The green curve corresponds to *k*-mers shared by both haplotypes, and the red and blue curves show *k*-mers found only in one of the haplotypes.

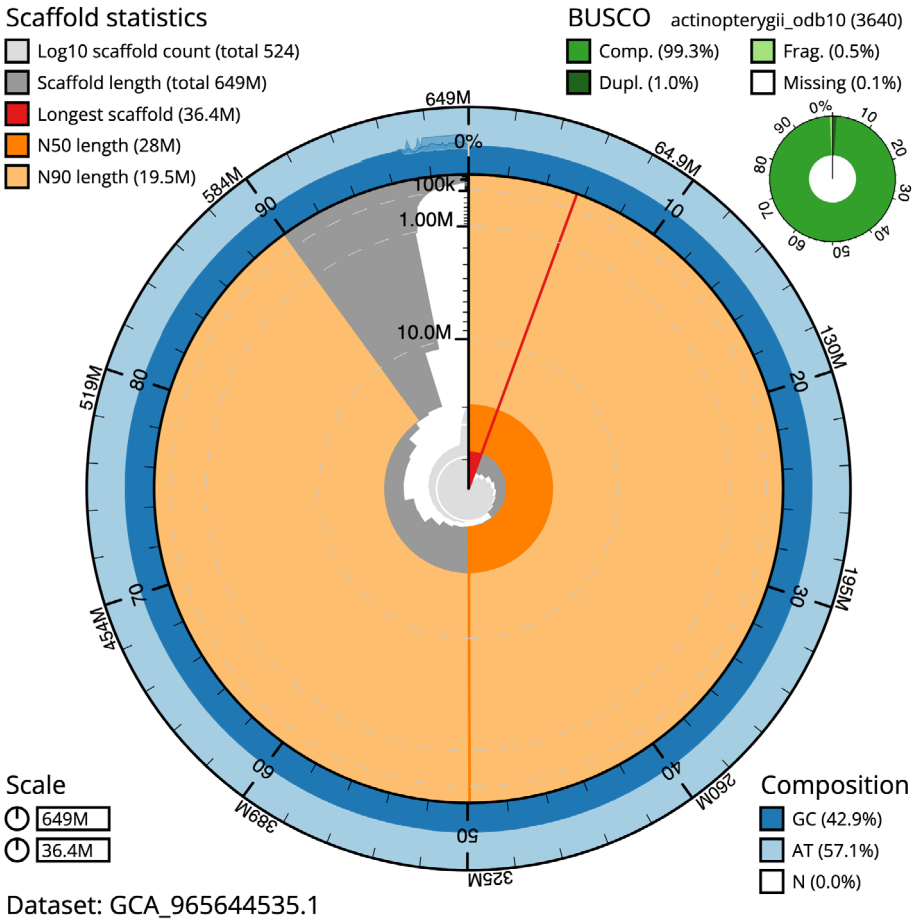

Figure 5. Assembly metrics for fCheLuc1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

which can be found in full on the Darwin Tree of Life website. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project. Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances, other Darwin Tree of Life collaborators.

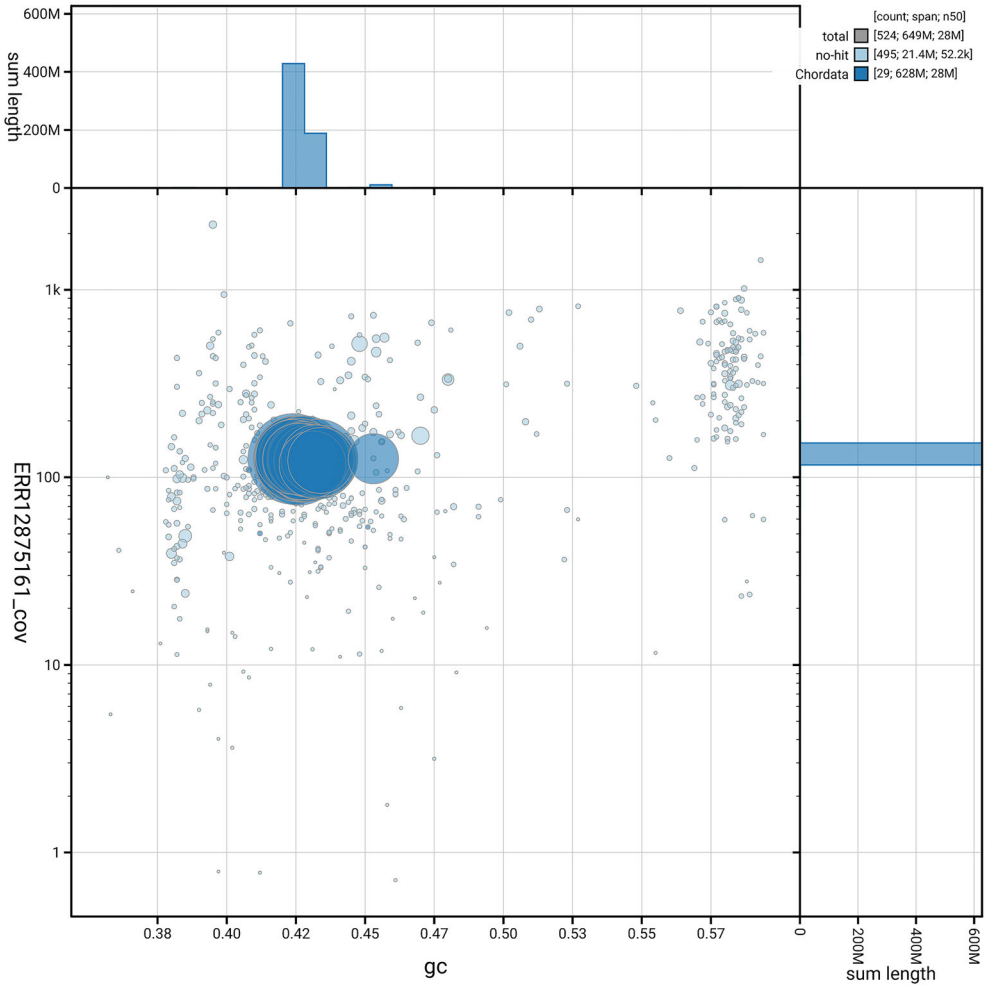

Figure 6. BlobToolKit GC-coverage plot for fCheLuc1.hap1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the BlobToolKit viewer.

Table 4. Earth Biogenome Project summary metrics for the *C. lucerna* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q61	6.C.Q40
Contig N50 length	3.03 Mb	≥ 1 Mb
Scaffold N50 length	27.99 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 61.4; haplotype 2: 62.1; combined: 61.7	≥ 40
k-mer completeness	Haplotype 1: 84.28%; Haplotype 2: 84.25%; combined: 99.62%	≥ 95%
BUSCO	C:99.3% [S:98.3%; D:1.0%]; F:0.5%; M:0.1%; n:3 640	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	96.66%	≥ 90%

Notes: The EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO: C = complete; S = single-copy; D = duplicated; F = fragmented; M = missing; n = orthologuesWellcome Sanger Institute – Legal and Governance.

Data availability

European Nucleotide Archive: Chelidonichthys lucerna (tub gurnard). Accession number PRJEB74613. The genome sequence is released openly for reuse. The *Chelidonichthys lucerna* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), the Sanger Institute Tree of Life Programme (PRJEB43745) and the Vertebrate Genomes Project (PRJNA489243). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

Production code used in genome assembly at the WSI Tree of Life is available at https://github.com/sanger-tol. Table 5 lists software versions used in this study.

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.4.6	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.8.3	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
samtools	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/ blobtoolkit	v0.8.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/ curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

Author information

Contributors are listed at the following links:

- Members of the Marine Biological Association Genome Acquisition Lab
- Members of the Darwin Tree of Life Barcoding collective
- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

References

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
Publisher Full Text

Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
Publisher Full Text

Baron J: **Les Triglidés (Teleostéens, Scorpaeniformes) de la Baie de Douarnenez. II. La reproduction de Eutrigla gurnardus, Trigla lucerna, Trigloparus lastoviza et Aspitrigla cuculus.** *Cybium.* 1985; **9**: 255–261.

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
Publisher Full Text

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
Publisher Full Text

Campos J, Costa- Dias S, Bio A, *et al.*: **Age and Growth of Tub Gurnard Chelidonichthys lucerna (Linnaeus, 1758) during Estuarine Occupation of a Temperate Atlantic Nursery.** *Int J Environ Sci Nat Resour.* 2022, **31**. [Google Scholar] [CrossRef]

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
Publisher Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
Publisher Full Text

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
Publisher Full Text

Colloca F, Cardinale M, Marcello A, *et al.*: **Tracing the life history of red gurnard (*Aspitrigla cuculus*) using validated otolith annual rings.** *J Appl Ichthyol.* 2003; **19**(1): 1–9.
Publisher Full Text

Colloca F, Milisenda G, Capezzuto F, *et al.*: **Spatial and temporal trend in the abundance and distribution of gurnards (Pisces: Triglidae) in the northern Mediterranean Sea.** *SciMar.* 2019; **83**: 101–116.
Publisher Full Text

Coull KA, Jermyn AS, Newton AW, *et al.*:Length/weight relationships for 88 species of fish encountered in the northeast Atlantic. Department of Agriculture and Fisheries for Scotland, Scottish Fisheries Research Report Number, 43, 1989; 82 pp.

Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the British terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123.
PubMed Abstract | Publisher Full Text | Free Full Text

Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *Gigascience.* 2021; **10**(2): giab008.
Publisher Full Text

Darwin Tree of Life Project Consortium: **Sequence locally, think globally: the Darwin Tree of Life Project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
Publisher Full Text

Deval MC, Bök T, Ates C, *et al.*: **Size selectivity of three diamond mesh codends for the European hake (Merluccius merluccius) and the tub gurnard (Trigla lucerna) in the Sea of Marmara, Turkey.** *J Appl Ichthyol.* 2007; **23**(2): 167–172.

El-Serafy SS, El-Gammal FI, Mehanna SF, *et al.*: **Age, Growth and Reproduction of the Tub Gurnard, Chelidonichthys lucerna (Linnaeus, 1758) from the Egyptian Mediterranean Waters off, Alexandria.** *Int J Fish Aquat Sci.* 2015; **4**: 13–20.
Publisher Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | Publisher Full Text | Free Full Text

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | Publisher Full Text

Forest A: Ressources halieutiques hors quotas du Nord Est Atlantique: Bilan des connaissances et analyse de scénarios d'évolution de la gestion. 2001.
Reference Source

Formenti G, Abueg L, Brajuka A, *et al.*: **Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
PubMed Abstract | Publisher Full Text | Free Full Text

Froese R, Pauly D. Editors: FishBase. World Wide Web electronic publication. 2017. version (06/2017).
Reference Source

Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
Publisher Full Text

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *Gigascience.* 2021; **10**(1): 33420778.
Publisher Full Text

ICES: Report of the Working Group on Assessment of New MOU Species (WGNEW). ICES CM 2007/ACFM 01. 2007.

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
Publisher Full Text

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
Publisher Full Text

Lawniczak MKN, Davey RP, Rajan J, *et al.*: **Specimen and sample metadata standards for biodiversity genomics: a proposal from the Darwin Tree of Life project [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2022; **7**: 187.
Publisher Full Text

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Lopez-Lopez L, Preciado I, Velasco F, *et al.*: **Resource partitioning among five coexisting species of gurnards (Scorpaeniforme: Triglidae): role of trophic and habitat segregation.** *J Sea Res.* 2011; **66**(2): 58–68.

Manni M, Berkeley MR, Seppey 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**(10): 4647–4654.
PubMed Abstract | **Publisher Full Text** 8476166

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
Reference Source

O'Leary NA, Cox E, Holmes JB, *et al.*: **Exploring and retrieving sequence and metadata for species across the Tree of Life with NCBI datasets.** *Sci Data.* 2024; **11**(1): 732.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Papaconstantinou C: **Age and growth of the yellow gurnard (*Trigla lucerna* L.1758) from the Thermaikos gulf (Greece) with some comments on its biology.** *Fish Res.* 1984; **2**(4): 243–255.
Publisher Full Text

ParkerHumphreys M: **Distribution and relative abundance of demersal fishes from beam trawl surveys in eastern English Channel (ICES division VIId) and the southern North Sea (ICES division IVc) 19932001.** *Science Series Technical Reports, CEFAS Lowestoft.* 2005; **124**: 92.
Reference Source

ParkerHumphreys M: **Distribution and relative abundance of demersal fishes from beam trawl surveys in the Irish Sea (ICES division VIIa) 19932001.** *Science Series Technical Reports, CEFAS Lowestoft.* 2004; **120**, 68. pp. 377.

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun.* 2020; **11**(1): 1432.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
Publisher Full Text

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Riedl R: *Fauna e flora del Mediterraneo.* Padova: Franco Muzzio Editore; 1991.

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI Taxonomy: a comprehensive update on curation, resources and tools.** *Database (Oxford)* 2020; **2020**: baaae062.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

SSP R, Huntley HM, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
Publisher Full Text

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2024; **9**: 339.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
Publisher Full Text

Vallisneri M, Stagioni M, Montanini S, *et al.*: **Body size, sexual maturity and diet in Chelidonichthys lucerna (Osteichthyes: Triglidae) from the Adriatic Sea, north eastern Mediterranean.** *Acta Adriat.* 2011; **51**(1): 141–148.

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems**. In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 20 March 2026

https://doi.org/10.21956/wellcomeopenres.28685.r151043

✔ **Bingpeng Xing**
Third Institute of Oceanography Ministry of Natural Resources, Xiamen, Fujian, China

The revised manuscript has satisfactorily addressed the major issues raised in the previous round, particularly the incorrect species information in the Background section, which has now been corrected. The genome assembly remains of high quality, and the overall manuscript has improved substantially. A few minor issues still remain, including an inconsistency in the species name spelling (*Chelidonichthys lucerne*?), insufficiently explicit explanation of k-mer completeness, and the need for clarification of how sequencing coverage was calculated. These are all minor textual issues.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: marine biodiversity；DNA barcoding；genome；marine ecosystems；mangrove

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.

Version 1

Reviewer Report 29 January 2026

https://doi.org/10.21956/wellcomeopenres.28003.r144346

Ivan Rodrigo Wolf
UNC Charlotte Department of Biological Sciences, Charlotte, North Carolina, USA

The manuscript "The genome sequence of the red gurnard, *Chelidonichthys lucerna* (Linnaeus, 1758) (Perciformes: Triglidae)" describes the genome assembly of *Chelidonichthys lucerna*. This fish is widely distributed in the eastern Atlantic and adjacent seas (Mediterranean and Black Sea), inhabiting depths ranging from 20 to 250 meters.

As part of Darwin Tree of Life Project, the methods follow a series of standardized protocols for DNA and RNA extraction and preparation for sequencing at the Wellcome Sanger Institute (WSI). The specimen used to obtain biomolecules was identified by a specialist and confirmed by DNA barcoding. In addition, the bioinformatics protocols utilize a set of current and classic softwares for the assembly and curation of *C. lucerna* haplotypes.

The assembly quality metrics are excellent compared to the benchmarks, with only the completeness of the *k*-mers being below expectations. Unfortunately, the data still lacks robust annotations, but the RNA-Seq data is publicly available.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, systems biology, genomics, transcriptomics, molecular phylogenetics, immunogenomics, pattern recognition, search and characterization of genes of biotechnological interest.

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 13 January 2026

https://doi.org/10.21956/wellcomeopenres.28003.r144349

Bingpeng Xing
Third Institute of Oceanography Ministry of Natural Resources, Xiamen, Fujian, China

This manuscript presents a high-quality chromosome-level genome assembly of Chelidonichthys lucerna generated using state-of-the-art PacBio HiFi and Hi-C technologies. The assembly metrics meet or exceed Earth BioGenome Project reference standards.

However, 1)there is a critical species-name inconsistency in the Background section (*C. cuculus* vs *C. lucerna*), and several areas would benefit from clearer explanation of haplotype completeness and assembly interpretation.

2)The full binomial (*Chelidonichthys lucerna*) should be used at its first occurrence, and the abbreviated form (*C. lucerna*) may be used thereafter.

3)Explicitly state the intended use of haplotype 2 (for phasing completeness rather than downstream analyses). Clarify that reduced k-mer completeness for individual haplotypes is an expected outcome of haplotype-resolved assembly, not an indicator of reduced assembly quality.

4)The reported ~127× genome coverage is appropriate but it is not entirely clear whether this refers to raw HiFi read coverage or post-filtering effective coverage.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Yes

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: marine biodiversity；DNA barcoding；genome；marine ecosystems；mangrove

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.