Check for updates

DATA NOTE

The genome sequence of the bronze bream, *Abramis brama* (Linnaeus, 1758) (Cypriniformes: Leuciscidae)

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

Andy D. Nunn[1], Bernd Hänfling[2], Richard Pitman[3],
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]Hull International Fisheries Institute, School of Natural Sciences, University of Hull, Hull, England, UK
[2]Institute for Biodiversity and Freshwater Conservation, University of the Highlands and Islands, Inverness, Scotland, UK
[3]Environment Agency, Bristol, England, UK

v1 **First published:** 28 Jul 2025, **10**:371
https://doi.org/10.12688/wellcomeopenres.24606.1
Latest published: 28 Jul 2025, **10**:371
https://doi.org/10.12688/wellcomeopenres.24606.1

Abstract

We present a genome assembly from a specimen of *Abramis brama* (bronze bream; Chordata; Actinopteri; Cypriniformes; Leuciscidae). The genome sequence has a total length of 1 108.20 megabases. Most of the assembly (99.13%) is scaffolded into 25 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 16.61 kilobases. Gene annotation of this assembly on Ensembl identified 25 268 protein-coding genes.

Keywords

Abramis brama, bronze bream, genome sequence, chromosomal, Cypriniformes

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ? ?

	1	2	3
version 1 28 Jul 2025	✓ view	? view	? view

1. **Jayan Duminda M Senevirathna**, Uva Wellassa University, Badulla, Sri Lanka
2. **Qiong Shi**, Shenzhen University, Shenzhen, China
3. **Karim Gharbi**, Earlham Institute, Norwich, UK

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: Nunn AD: Writing – Original Draft Preparation; **Hänfling B**: Investigation, Resources, Writing – Review & Editing; **Pitman R** : Investigation, Resources;

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.

Copyright: © 2025 Nunn AD *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: Nunn AD, Hänfling B, Pitman R *et al.* **The genome sequence of the bronze bream, *Abramis brama* (Linnaeus, 1758) (Cypriniformes: Leuciscidae) [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2025, **10**:371 https://doi.org/10.12688/wellcomeopenres.24606.1

First published: 28 Jul 2025, **10**:371 https://doi.org/10.12688/wellcomeopenres.24606.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Actinopterygii; Actinopteri; Neopterygii; Teleostei; Osteoglossocephalai; Clupeocephala; Otomorpha; Ostariophysi; Otophysi; Cypriniphysae; Cypriniformes; Cyprinoidei; Leuciscidae; Leuciscinae; *Abramis*; *Abramis brama* (Linnaeus, 1758) (NCBI:txid38527)

Background

The common bream *Abramis brama* (L.) is a deep-bodied, laterally compressed fish that inhabits lowland fresh waters, and occasionally estuarine and brackish environments (Kafemann *et al.*, 2000), across most of Europe and parts of western and central Asia (Kottelat & Freyhof, 2007). Individuals attain sexually maturity at 3 to 5 years of age, have an average life span of 6 to 12 years and can grow moderately large (typically to 30–50 cm in length) in suitable conditions (Maitland, 2004). For identification purposes, a useful feature is the long-based anal fin, which has 24 to 30 fully branched rays (Maitland, 2004).

Common bream often form large shoals and are invariably most abundant in slow-flowing rivers, flood plains and productive lakes and reservoirs. Habitat use is relatively flexible, but distinct temporal and ontogenetic shifts, such as between rivers and flood plains on a seasonal basis and from shallow, vegetated areas to deeper water during development, sometimes occur (Borcherding *et al.*, 2002; Grift *et al.*, 2001; Stoffers *et al.*, 2025). Adults can be highly mobile, with some undertaking long-distance migrations, although inter-individual variability is often considerable (Brodersen *et al.*, 2019; Gardner *et al.*, 2013; Winter *et al.*, 2021). Spawning occurs in single or multiple events in late spring or early summer (Nunn *et al.*, 2007a; Poncin *et al.*, 1996), with females each depositing ~100 000 to 600 000 eggs on a variety of surfaces in shallow water (Maitland, 2004). Larvae and juveniles forage largely on zooplankton before becoming benthivorous (Nunn *et al.*, 2007b; Persson & Brönmark, 2002). The benthic foraging behaviour of adults can involve significant bioturbation and, thus, lead to increases in turbidity and nutrient mobilisation, as well as reduced light penetration and damage to submerged vegetation, which can affect food-web structure and ecosystem functioning (Carvalho & Moss, 1995; Manni *et al.*, 2021; Volta *et al.*, 2013).

As a consequence of its wide geographical range and large and stable populations, the common bream is currently listed as "Least Concern" on the International Union for the Conservation of Nature (IUCN) Red List of Threatened Species (Ford, 2024). Notwithstanding, the species is impacted by chronic habitat degradation and fragmentation, and some populations are threatened by overharvesting (Ford, 2024). Additionally, common bream are prone to hybridising with closely related species, particularly in heavily modified habitats, which has implications for the genetic integrity of the populations (Cowx, 1983; Hayden *et al.*, 2010; Konopiński & Amirowicz, 2018).

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 – we present the first reference genome sequence for *Abramis brama*, the bronze bream. This genome was assembled using the Tree of Life pipeline from a specimen collected in Calverton Fish Farm, Lake Side View, Moor Ln, Calverton, Nottingham, United Kingdom (Figure 1).

Methods

Sample acquisition

One juvenile *A. brama* (specimen ID SAN0000702, ToLID fAbrBra2; Figure 1), was collected from the Environment Agency's National Coarse Fish Rearing Unit in Calverton, Nottingham, UK (latitude 53.03, longitude –1.05) on 25 August 2020. The National Coarse Fish Rearing Unit at Calverton is funded solely through rod licence duty. The specimen was taken from the nursery pond by Richard Pitman (Environment Agency) using a seine net, and left to recover fully in fresh flowing, clean, borehole water for a week before any sampling commenced. The specimen was transported alive to the University of Hull where they were identified by Bernd Hänfling (University of Hull) and euthanised in a lethal dose of MS-222. Tissues dissection took place within 30 minutes of euthanasia, and the tissues were immediately shock-frozen in liquid nitrogen. Sample metadata was collected in line with the Darwin Tree of Life project standards described by Lawniczak *et al.* (2022).

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 fAbrBra2 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from

Figure 1. Photograph of the *Abramis brama* (fAbrBra2) specimen from which samples were taken for genome sequencing.

the gill animal was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted 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 automated 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 17.3 ng/μL and a yield of 2 249.00 ng. The 260/280 spectrophotometric ratio was 1.81, and the 260/230 ratio was 1.93.

RNA was extracted from spleen tissue of fAbrBra2 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 using the Sequel IIe system (Pacific Biosciences, California, USA). The concentration of the library loaded onto the Sequel IIe was in the range 40–135 pM. The SMRT link software, a PacBio web-based end-to-end workflow manager, was used to set-up and monitor the run, and to perform primary and secondary analysis of the data upon completion.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen muscle tissue of the fAbrBra2 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 and 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 6000.

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 6000 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 (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with 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 18 breaks, 27 joins, and removal of 33 haplotypic duplications. 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 the primary and alternate haplotypes using the *k*-mer databases (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 by querying the GoaT database (Challis *et al.*, 2023). 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 package management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *Abramis brama* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 48.15 Gb (gigabases) from 4.55 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 060.72 Mb, with a heterozygosity of 0.55% and repeat content of 35.95% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 44× coverage. Hi-C sequencing produced 137.55 Gb from 910.94 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.

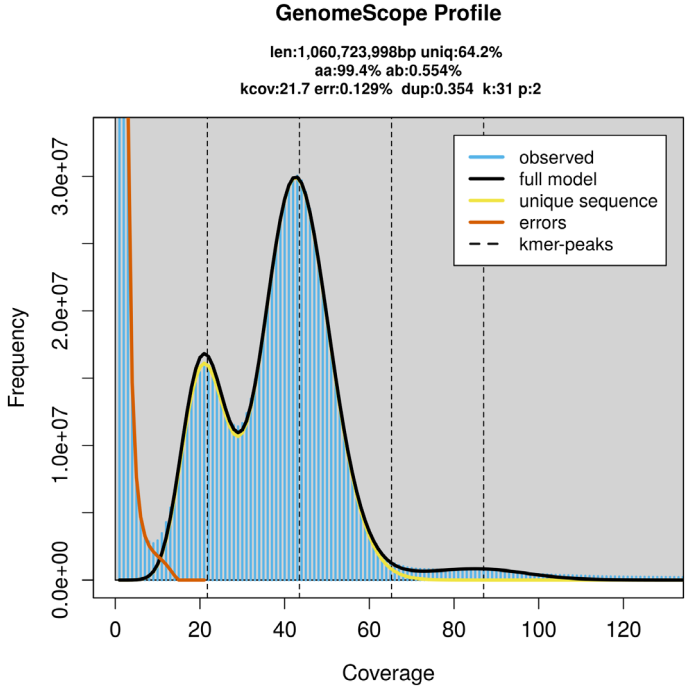

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB71619.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	fAbrBra2	fAbrBra2	fAbrBra2
Specimen ID	SAN0000702	SAN0000702	SAN0000702
BioSample (source individual)	SAMEA11296536	SAMEA11296536	SAMEA11296536
BioSample (tissue)	SAMEA11296574	SAMEA11296576	SAMEA11296572
Tissue	gill animal	muscle	spleen
Sequencing platform and model	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR12408800; ERR12408799	ERR12512746	ERR12512747
Read count total	4.55 million	910.94 million	58.42 million
Base count total	48.15 Gb	137.55 Gb	8.82 Gb

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The final assembly has a total length of 1 108.20 Mb in 207 scaffolds, with 685 gaps, and a scaffold N50 of 42.79 Mb (Table 2). Most of the assembly sequence (99.13%) was assigned to 25 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.

Table 2. Genome assembly statistics.

Assembly name	fAbrBra2.1
Assembly accession	GCA_963993115.1
Alternate haplotype accession	GCA_963993125.1
Assembly level	chromosome
Span (Mb)	1 108.20
Number of chromosomes	25
Number of contigs	892
Contig N50	3.48 Mb
Number of scaffolds	207
Scaffold N50	42.79 Mb
Sex chromosomes	N/A
Organelles	Mitochondrial genome: 16.61 kb

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 59.2. The *k*-mer completeness is 88.21% for the primary assembly, 83.65% for the alternate haplotype, and 99.11% for the combined assemblies (Figure 4). BUSCO v.5.5.0 analysis using the actinopterygii_odb10 reference set (n = 3 640) identified 3559% of the expected gene set (single = 3496%, duplicated = 63%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the primary assembly, is **6.C.Q59**, meeting the recommended reference standard.

Genome annotation report

The *Abramis brama* genome assembly (GCA_963993115.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 86 080 transcribed mRNAs from 25 268 protein-coding and 14 948 non-coding genes. The average transcript length is 23 408.71 bp, with an average of 2.14 coding transcripts per gene and 11.86 exons per transcript. For further information about the annotation, please refer to the annotation page on Ensembl.

Wellcome Sanger Institute – Legal and Governance

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'**, 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

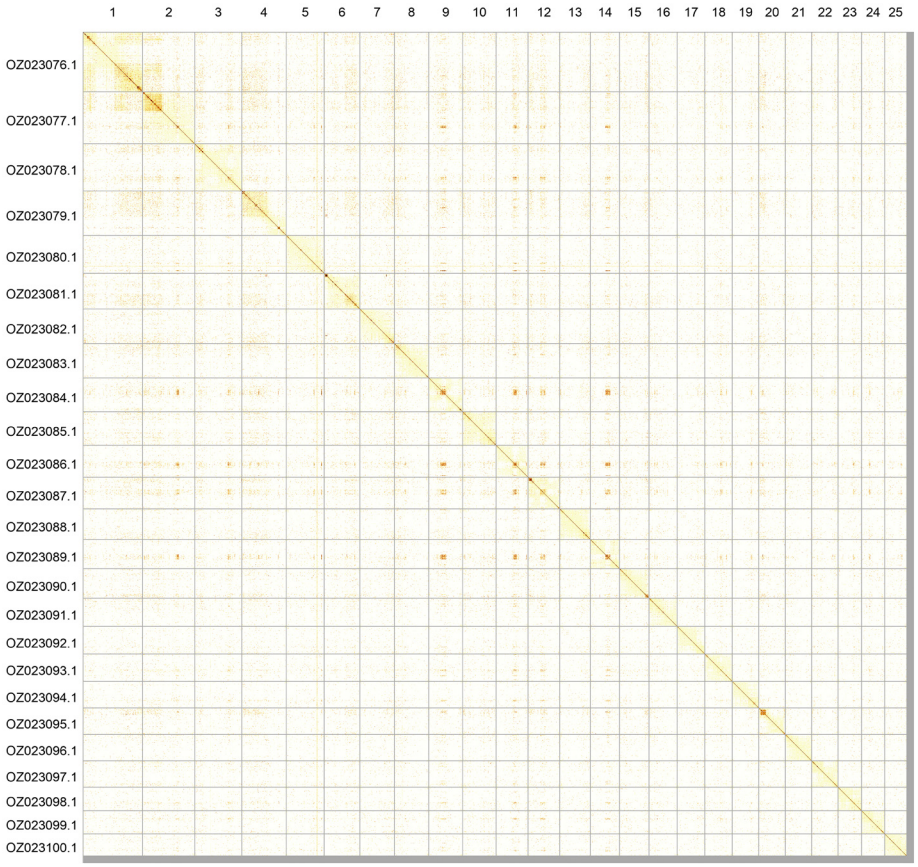

Figure 3. Hi-C contact map of the *Abramis brama* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Abramis brama* fAbrBra2.

INSDC accession	Molecule	Length (Mb)	GC%
OZ023076.1	1	79.94	39
OZ023077.1	2	69.26	38.50
OZ023078.1	3	62.81	38.50
OZ023079.1	4	59.17	39
OZ023080.1	5	50.80	38.50
OZ023081.1	6	47.36	39
OZ023082.1	7	46.18	38.50
OZ023083.1	8	45.73	38.50
OZ023084.1	9	45.25	39
OZ023085.1	10	44.55	39
OZ023086.1	11	42.79	38.50
OZ023087.1	12	42.26	38.50
OZ023088.1	13	40.40	39
OZ023089.1	14	39.32	38.50
OZ023090.1	15	39.15	38.50
OZ023091.1	16	37.72	38.50
OZ023092.1	17	36.76	38.50
OZ023093.1	18	36.41	38.50
OZ023094.1	19	35.60	38.50
OZ023095.1	20	35.29	39
OZ023096.1	21	35.27	38
OZ023097.1	22	35.05	38.50
OZ023098.1	23	31.40	38.50
OZ023099.1	24	30.63	38.50
OZ023100.1	25	29.42	38.50
OZ023101.1	MT	0.02	43

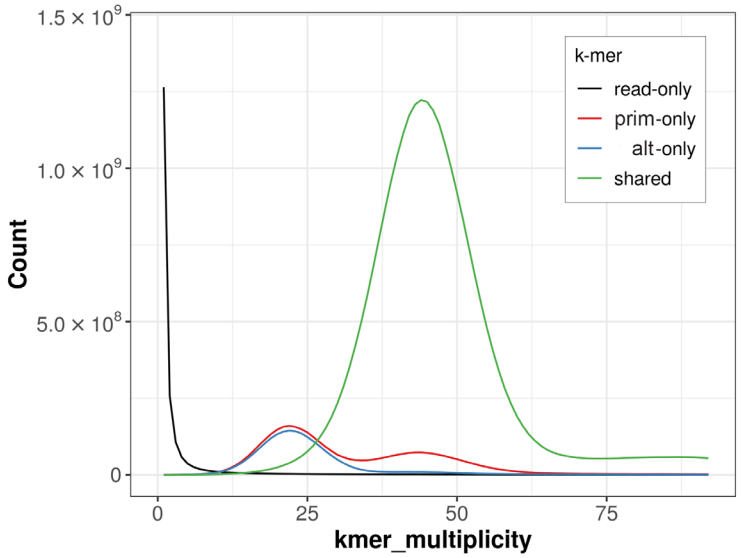

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

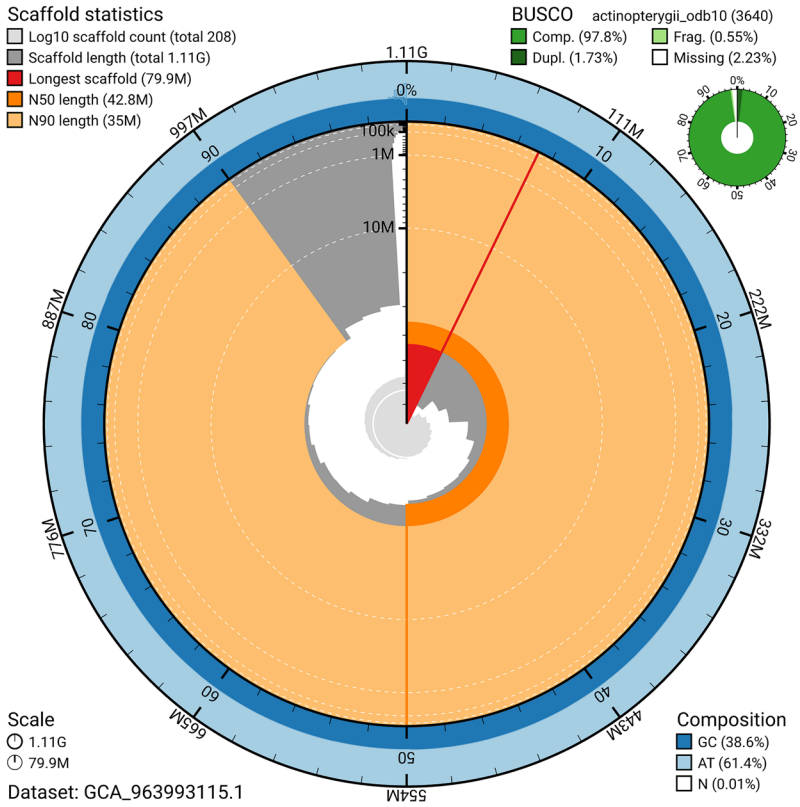

Figure 5. Assembly metrics for fAbrBra2.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue tracks 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 actinopterygii_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

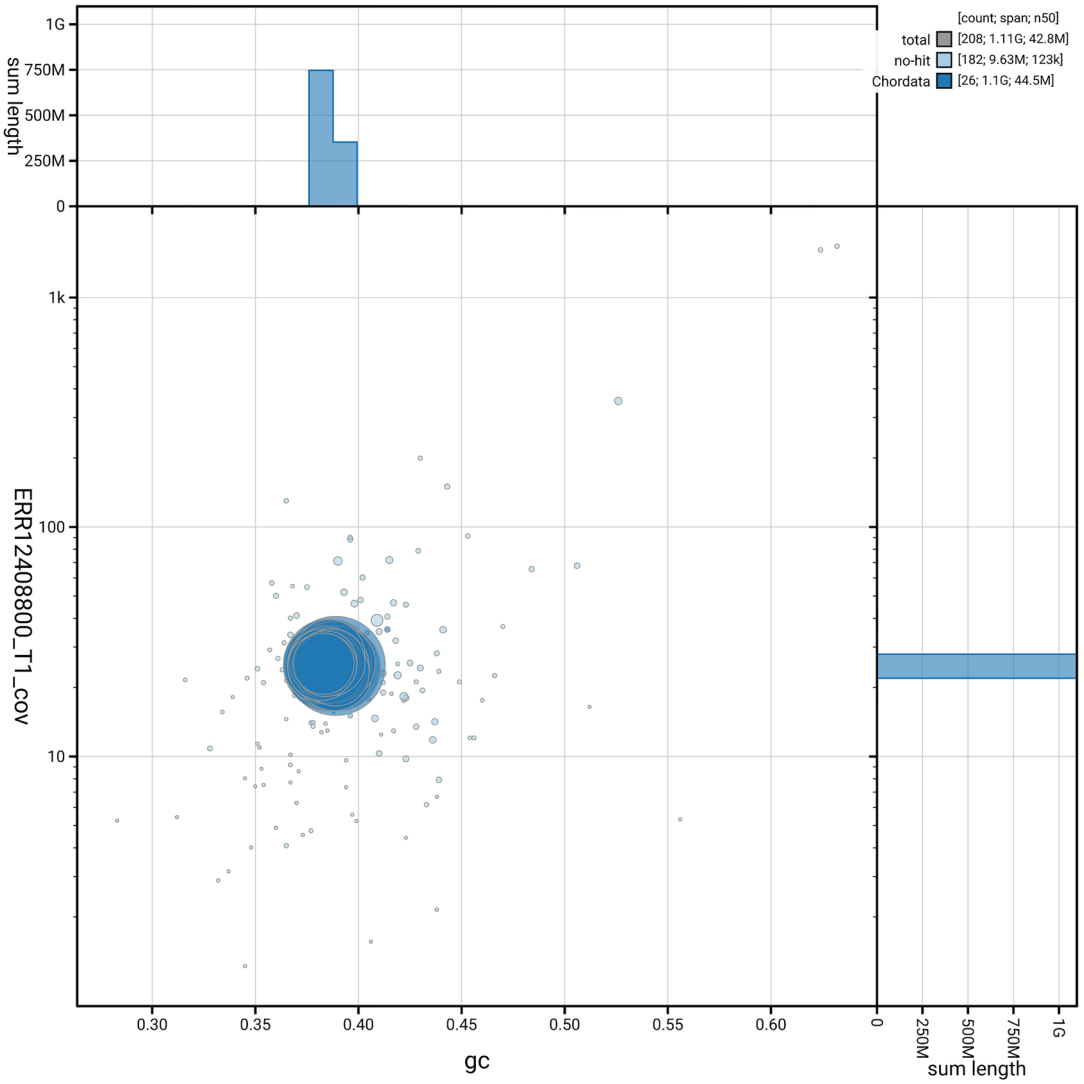

Figure 6. BlobToolKit GC-coverage plot for fAbrBra2.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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 *Abramis brama* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q59	6.C.Q40
Contig N50 length	3.48 Mb	≥ 1 Mb
Scaffold N50 length	42.79 Mb	= chromosome N50
Consensus quality (QV)	Primary: 59.7; alternate: 58.8; combined: 59.2	≥ 40
k-mer completeness	Primary: 88.21%; alternate: 83.65%; combined: 99.11%	≥ 95%
BUSCO	C:97.8%[S:96.0%;D:1.7%]; F:0.5%; M:1.7%; n:3 640	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.13%	≥ 90%

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.

Data availability

European Nucleotide Archive: Abramis brama (common bream). Accession number PRJEB71619. The genome sequence is released openly for reuse. The *Abramis brama* 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. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

Pipelines used for genome assembly at the WSI Tree of Life are available at https://pipelines.tol.sanger.ac.uk/pipelines. Table 5 lists software versions used in this study.

Author information

- 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

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	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	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	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi

Software	Version	Source
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.04.1	https://github.com/nextflow-io/nextflow
PretextSnapshot		https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.4.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the universal protein knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Borcherding J, Bauerfeld M, Hintzen DD, *et al.*: **Lateral migrations of fishes between floodplain lakes and their drainage channels at the Lower Rhine: diel and seasonal aspects.** *J Fish Biol.* 2002; **61**(5): 1154–1170.
Publisher Full Text

Brodersen J, Hansen JH, Skov C: **Partial nomadism in large-bodied bream (*Abramis brama*).** *Ecol Freshw Fish.* 2019; **28**(4): 650–660.
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.
PubMed Abstract | Publisher Full Text | Free Full Text

Carvalho L, Moss B: **The current status of a sample of english sites of special scientific interest subject to eutrophication.** *Aquat Conserv.* 1995; **5**(3): 191–204.
Publisher Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Cowx IG: **The biology of bream, *Abramis brama* (L.), and its natural hybrid with roach, *Rutilus rutilus* (L.), In the river exe.** *J Fish Biol.* 1983; **22**(6): 631–646.
Publisher Full Text

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

Ford M: **Abramis brama.** *The IUCN Red List of Threatened Species.* 2024.
Publisher Full Text

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

Gardner CJ, Deeming DC, Eady PE: **Seasonal movements with shifts in lateral and longitudinal habitat use by common bream, *Abramis brama*, in a heavily modified lowland river.** *Fish Manag Ecol.* 2013; **20**(4): 315–325.
Publisher Full Text

Grift RE, Buijse AD, Klein Breteler JPG, *et al.*: **Migration of bream between the main channel and floodplain lakes along the lower River Rhine during the connection phase.** *J Fish Biol.* 2001; **59**(4): 1033–1055.
Publisher Full Text

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476.
PubMed Abstract | Publisher Full Text | Free Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | Publisher Full Text | Free Full Text

Hayden B, Pulcini D, Kelly-Quinn M, *et al.*: **Hybridisation between two cyprinid fishes in a novel habitat: genetics, morphology and life-history traits.** *BMC Evol Biol.* 2010; **10**: 169.
PubMed Abstract | Publisher Full Text | Free Full Text

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): giaa153.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kafemann R, Adlerstein S, Neukamm R: **Variation in otolith strontium and calcium ratios as an indicator of life-history strategies of freshwater fish species within a brackish water system.** *Fish Res.* 2000; **46**(1–3): 313–325.
Publisher Full Text

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.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Konopiński MK, Amirowicz A: **Genetic composition of a population of natural common bream *Abramis brama* × roach *Rutilus rutilus* hybrids and their morphological characteristics in comparison with parent species.** *J Fish Biol.* 2018; **92**(2): 365–385.
PubMed Abstract | **Publisher Full Text**

Kottelat M, Freyhof J: **Handbook of European freshwater fishes.** Berlin: Publications Kottelat, 2007.
Reference Source

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

Maitland P: **Keys to the freshwater fish of Great Britain and Ireland with notes on their distribution and ecology.** Ambleside: Freshwater Biological Association, 2004.
Reference Source

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** | **Free Full Text**

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2, [Accessed 2 April 2024].
Reference Source

Nunn AD, Harvey JP, Cowx IG: **Variations in the spawning periodicity of eight fish species in three english lowland rivers over a 6 year period, inferred from 0+ year fish length distributions.** *J Fish Biol.* 2007a; **70**(4): 1254–1267.
Publisher Full Text

Nunn AD, Harvey JP, Cowx IG: **The food and feeding relationships of larval and 0+ year juvenile fishes in lowland rivers and connected waterbodies. I. Ontogenetic shifts and interspecific diet similarity.** *J Fish Biol.* 2007b; **70**(3): 726–742.
Publisher Full Text

Persson A, Brönmark C: **Foraging capacities and effects of competitive release on ontogenetic diet shift in bream, *Abramis brama*.** *Oikos.* 2002; **97**(2): 271–281.
Publisher Full Text

Poncin P, Philippart JC, Ruwet JC: **Territorial and non-territorial spawning behaviour in the bream.** *J Fish Bio.* 1996; **49**(4): 622–26.
Publisher 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**(1): 1432.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rao SSP, Huntley MH, 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.
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.
PubMed Abstract | **Publisher Full Text** | **Free 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**

Stoffers T, Buijse AD, Poos JJ, *et al.*: **Ontogenetic shifts by juvenile fishes highlight the need for habitat heterogeneity and connectivity in river restoration.** *Limnol Oceanogr.* 2025; **70**(3): 732–748.
Publisher 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.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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

Volta P, Jeppesen E, Leoni B, *et al.*: **Recent invasion by a non-native cyprinid (common bream *Abramis brama*) is followed by major changes in the ecological quality of a shallow lake in southern Europe.** *Biol Invasions.* 2013; **15**: 2065–2079.
Publisher Full Text

Winter ER, Hindes AM, Lane S, *et al.*: **Movements of common bream *Abramis brama* in a highly connected, lowland wetland reveal sub-populations with diverse migration strategies.** *Freshw Biol.* 2021; **66**(7): 1410–1422.
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 1

Reviewer Report 16 September 2025

https://doi.org/10.21956/wellcomeopenres.27116.r130958

© **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
Earlham Institute, Norwich, England, UK

This data note describes a high-quality reference assembly of the common bream (Abramis brama) genome, with Ensembl annotation generated as part of the Darwin Tree of Life Project. The note includes a general overview of the species biology followed by a standardised genome report. The introduction is informative and the methods comprehensive. Assembly statistics and assessment are consistent with a high-quality, reference-grade genome assembled to chromosome-level scaffolds/pseudo-molecules. Overall, this study represents an important contribution to biodiversity genomics.

While not critical, the report contains a number of omissions and inconsistencies, which the authors may want to rectify:

1. The species is referred to both 'bronze' and 'common' bream throughout the text. I'd recommend picking one to avoid confusion.

2. The sex of the sampled individual is not described. If this is known, it should be added. If it's unknown, it should be stated.

3. The tissue for HMW extraction is described as 'gill animal'. This should probably read 'gill' or 'animal's gill'.

4. Kits and instruments are described differently between sections with regard to the manufacturer. This should be standardised, ideally.

5. Yield and purity of the RNA extraction is not provided, while it is for HMW DNA.

6. The number of PCR cycles for HiC library preparation is given as a range. This should be provided as a fixed value for this sample.

7. The sequencing chemistry used for Illumina sequencing is missing.

8. Basecelling software versions for Sequel IIe and NovaSeq are missing.

9. BUSCO metrics in the text, including % identified, % single, and % duplicated, seem to be incorrect and do not align with those in Figure 5.

10. The average transcript length is one order of magnitude larger than the average calculated from the corresponding fasta file in Ensembl.

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: Genomics and 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 02 September 2025

https://doi.org/10.21956/wellcomeopenres.27116.r130956

© 2025 Shi Q. 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.

?

Qiong Shi
College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, Guangdong, China

The authors performed whole-genome sequencing and constructed a chromosome-level genome assembly for the bronze bream. The mitochondrial genome of this fish was also assembled. Generally speaking, the overall writing of this manuscript is ok, and numerous data were generated to support the main conclusions. However, major revisions are still required before acceptance for indexing.
1. It is recommended to change the title of this data note as follows:

A chromosome-level genome assembly of the bronze bream (*Abramis brama*).
2. Background:
(1) Lines 5-6: Change “sexually” to “sexual”.
(2) Please unify the English name of *Abramis brama* (common bream or bronze bream).
3. Methods:
(1) The first line of page 4: What is “gill animal”? It also appears in the Table 1.
(2) Why didn’t you use muscle for the PacBio HiFi sequencing (as the Hi-C)?
4. Results:
(1) Under “Assembly quality metrics” (page 6): In lines 4-7, some data are obviously wrong.
(2) Figure 3: Please adjust this figure, since the current version is unclear.
5. Data availability: Make sure to release all the genomics data before acceptance of this manuscript.

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?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Fish genomics and molecular breeding

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 02 September 2025

https://doi.org/10.21956/wellcomeopenres.27116.r129974

© 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 provides valuable insights into the genome of an important fish species. The study holds significant relevance and contributes to the broader understanding of

this species' biology.

The scientific name of the species should be italicized in accordance with standard scientific writing conventions.

It would be beneficial to include information on the species' sexual dimorphism and special adaptations to its habitat as part of the background. This will provide readers with a stronger biological context for the genomic study.

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: 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.