DATA NOTE

The genome sequence of the stone loach, *Barbatula barbatula* (Linnaeus, 1758) [version 1; peer review: 1 approved]

Bernd Hänfling[1], Andy D. Nunn[2], Paolo Moccetti[2],
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Institute for Biodiversity and Freshwater Conservation, University of Highlands and Islands, Inverness, Scotland, UK
[2]School of Natural Sciences, University of Hull, Hull, England, UK

v1 **First published:** 13 Nov 2023, **8**:518
https://doi.org/10.12688/wellcomeopenres.20293.1
Latest published: 13 Nov 2023, **8**:518
https://doi.org/10.12688/wellcomeopenres.20293.1

Abstract

We present a genome assembly from an individual female *Barbatula barbatula* (the stone loach; Chordata; Actinopteri; Cypriniformes; Nemacheilidae). The genome sequence is 617.6 megabases in span. Most of the assembly is scaffolded into 25 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 16.64 kilobases in length.

Keywords

Barbatula barbatula, stone loach, genome sequence, chromosomal, Cypriniformes

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓

	1
version 1 13 Nov 2023	✓ view

1. **Guoqing Lu**, University of Nebraska Omaha, Omaha, 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: Hänfling B: Investigation, Resources, Writing – Review & Editing; **Nunn AD**: Investigation, Resources, Writing – Original Draft Preparation; **Moccetti P**: 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 (206194) and the Darwin Tree of Life Discretionary Award (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: Hänfling B, Nunn AD, Moccetti P *et al.* **The genome sequence of the stone loach, *Barbatula barbatula* (Linnaeus, 1758) [version 1; peer review: 1 approved]** Wellcome Open Research 2023, **8**:518
https://doi.org/10.12688/wellcomeopenres.20293.1

First published: 13 Nov 2023, **8**:518 https://doi.org/10.12688/wellcomeopenres.20293.1

Species taxonomy

Eukaryota; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Actinopterygii; Actinopteri; Neopterygii; Teleostei; Osteoglossocephalai; Clupeocephala; Otomorpha; Ostariophysi; Otophysi; Cypriniphysae; Cypriniformes; Cobitoidei; Nemacheilidae; *Barbatula*; *Barbatula barbatula* (Linnaeus, 1758) (NCBI:txid135647).

Background

Stone loach *Barbatula barbatula* (L.) has a broad Palearctic distribution that includes large parts of Europe and Asia (Froese & Pauly, 2023; Kottelat & Freyhof, 2007a). It is resident in freshwater and small, rarely exceeding 12 cm in length, with an elongated body and three barbels on either side of the mouth (Maitland, 2004). Age- and size-at-maturity is typically 1–3 years and 8–12 cm, respectively, and maximum life span 4–6 years (Wheeler, 1992). There appear to be no known sub-species, notable forms or variants, but many populations are geographically isolated and genetically distinct (Barluenga & Meyer, 2005; Šedivá *et al.*, 2008), and molecular studies suggest that several species may actually exist under the same name (Freyhof, 2011).

Stone loach appears to have a preference for shallow, well-oxygenated water flowing over relatively coarse substrata, such as cobble and gravel, but can tolerate a range of physico-chemical conditions and sometimes occur in areas with mud or silt (Smyly, 1955). Typical habitats include riffle-pool sequences in the upper and middle reaches of rivers, although it can also occur in lowland rivers and connected still waters (Mann, 1971; Nunn *et al.*, 2007; Prenda *et al.*, 1997; Welton *et al.*, 1983). Some studies have reported seasonal or ontogenetic differences in microhabitat use, such as shifts to higher water velocities and larger substrata with increasing age, but many individuals occupy similar areas during all stages of development and throughout the year (Copp & Vilizzi, 2004; Kováč *et al.*, 1999; Zweimüller, 1995).

Spawning usually occurs between April and June, with fewer than 2000 to more than 27,000 eggs per female deposited on gravel, sand or aquatic vegetation (Mann, 1971; Skryabin, 1993; Smyly, 1955). The reproductive strategy varies according to latitude and environmental conditions, often occurring as a single, discrete event towards the northern extent of the species' range and in unproductive waters, and as multiple events over a protracted period elsewhere (Mann *et al.*, 1984; Mills & Eloranta, 1985; Vinyoles *et al.*, 2010). Stone loach is iteroparous, with most adults spawning at least twice during their life (Mills & Eloranta, 1985). In an aquarium, stone loach eggs hatched after an incubation period of 14–16 days at 12–16 °C, with the switch from endogenous to exogenous feeding occurring after 5–7 days (Smyly, 1955). Larvae, juveniles and adults are benthivorous, consuming mainly small crustaceans and insect larvae (Hyslop, 1982; Maitland, 1965; Smyly, 1955; Welton *et al.*, 1983). Growth in body length is fastest in the first year of life and slowest, often coinciding with delayed maturity and greater longevity, in unproductive waters at high latitudes (Mann *et al.*, 1984; Mills *et al.*, 1983; Mills & Eloranta, 1985). Stone loach is generally sedentary, although short movements and migrations sometimes occur, particularly at night and prior to the spawning period (MacKenzie & Greenberg, 1988; Nunn *et al.*, 2010; Smyly, 1955).

Stone loach is classified as "Least Concern" at global and European levels in terms of extinction risk (Freyhof, 2011; Freyhof & Brooks, 2011), but faces localised threats due to a range of pressures. For example, although tolerant of mild organic pollution (Prenda *et al.*, 1997; Smyly, 1955), the species is highly sensitive to low dissolved oxygen concentrations and heavy metals (Kottelat & Freyhof, 2007b), and, given its poor swimming ability, even small obstacles could represent total migration barriers and lead to habitat fragmentation (Tudorache *et al.*, 2008). The stone loach is also likely to face predation and competition threats from some non-native species, such as crayfish (Bubb *et al.*, 2009; Galib *et al.*, 2021; Guan & Wiles, 1997).

Genome sequence report

The genome was sequenced from one female *Barbatula barbatula* (Figure 1) collected from the River Wharfe, UK (53.91, –1.61). A total of 30-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 247 missing joins or mis-joins and removed 19 haplotypic duplications, reducing the assembly length by 1.53% and the scaffold number by 56.71%, and increasing the scaffold N50 by 7.4%.

The final assembly has a total length of 617.6 Mb in 99 sequence scaffolds with a scaffold N50 of 24.4 Mb (Table 1). The snailplot in Figure 2 provides a summary of the assembly

Figure 1. Photographs of the *Barbatula barbatula* (fBarBar1) specimen used for genome sequencing.

Table 1. Genome data for *Barbatula barbatula*, fBarBar1.1.

Project accession data		
Assembly identifier	fBarBar1.1	
Assembly release date	2022-10-07	
Species	*Barbatula barbatula*	
Specimen	fBarBar1	
NCBI taxonomy ID	135647	
BioProject	PRJEB55341	
BioSample ID	SAMEA11296540	
Isolate information	fBarBar1 fBarBar1	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	59.4	*≥ 50*
k-mer completeness	100%	*≥ 95%*
BUSCO**	C:96.2%[S:94.8%,D:1.4%], F:0.8%,M:3.0%,n:3,640	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.73%	*≥ 95%*
Sex chromosomes	-	*localised homologous pairs*
Organelles	Mitochondrial genome assembled	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR10077563	
Hi-C Illumina	ERR10084070	
PolyA RNA-Seq Illumina	ERR10890703	
Genome assembly		
Assembly accession	GCA_947034865.1	
Accession of alternate haplotype	GCA_947034905.1	
Span (Mb)	617.6	
Number of contigs	925	
Contig N50 length (Mb)	2.7	
Number of scaffolds	99	
Scaffold N50 length (Mb)	24.4	
Longest scaffold (Mb)	36.4	

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from (Rhie *et al.*, 2021).

** BUSCO scores based on the actinopterygii_odb10BUSCO set using v5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/fBarBar1.1/dataset/CAMQQH01/busco.

statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.73%) of the assembly sequence was assigned to 25 chromosomal-level scaffolds. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). A region of undetermined order and orientation was observed on Chromosome

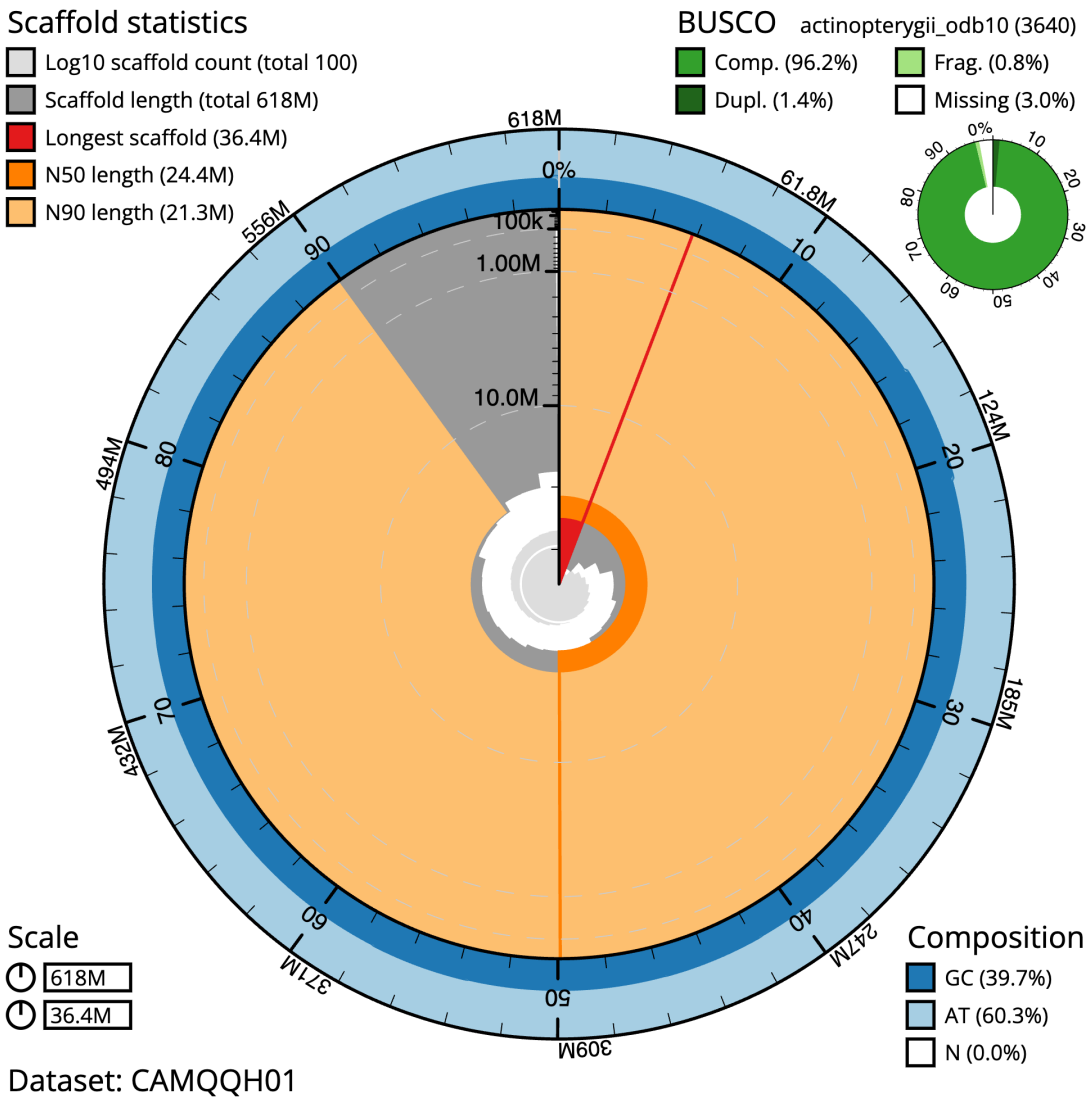

Figure 2. Genome assembly of *Barbatula barbatula*, fBarBar1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 617,663,352 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (36,424,156 bp, shown in red). . Orange and pale-orange arcs show the N50 and N90 scaffold lengths (24,423,978 and 21,267,045 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the actinopterygii_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Barbatula%20barbatula/dataset/CAMQQH01/snail.

4 in the region 12,000 kb to 23,500 kb. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 59.4 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 96.2% (single = 94.8%, duplicated = 1.4%), using the actinopterygii_odb10 reference set (n = 3,640).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/135647.

Methods

Sample acquisition and nucleic acid extraction

A female specimen of *B. barbatula* (specimen ID SAN0000706, ToLID fBarBar1) was collected from the River Wharfe, UK (latitude 53.91, longitude –1.61) on 2020-09-09. The specimen was collected by Andy Nunn and Paolo Moccetti using electro-fishing, and identified by Bernd Hänfling. The specimen

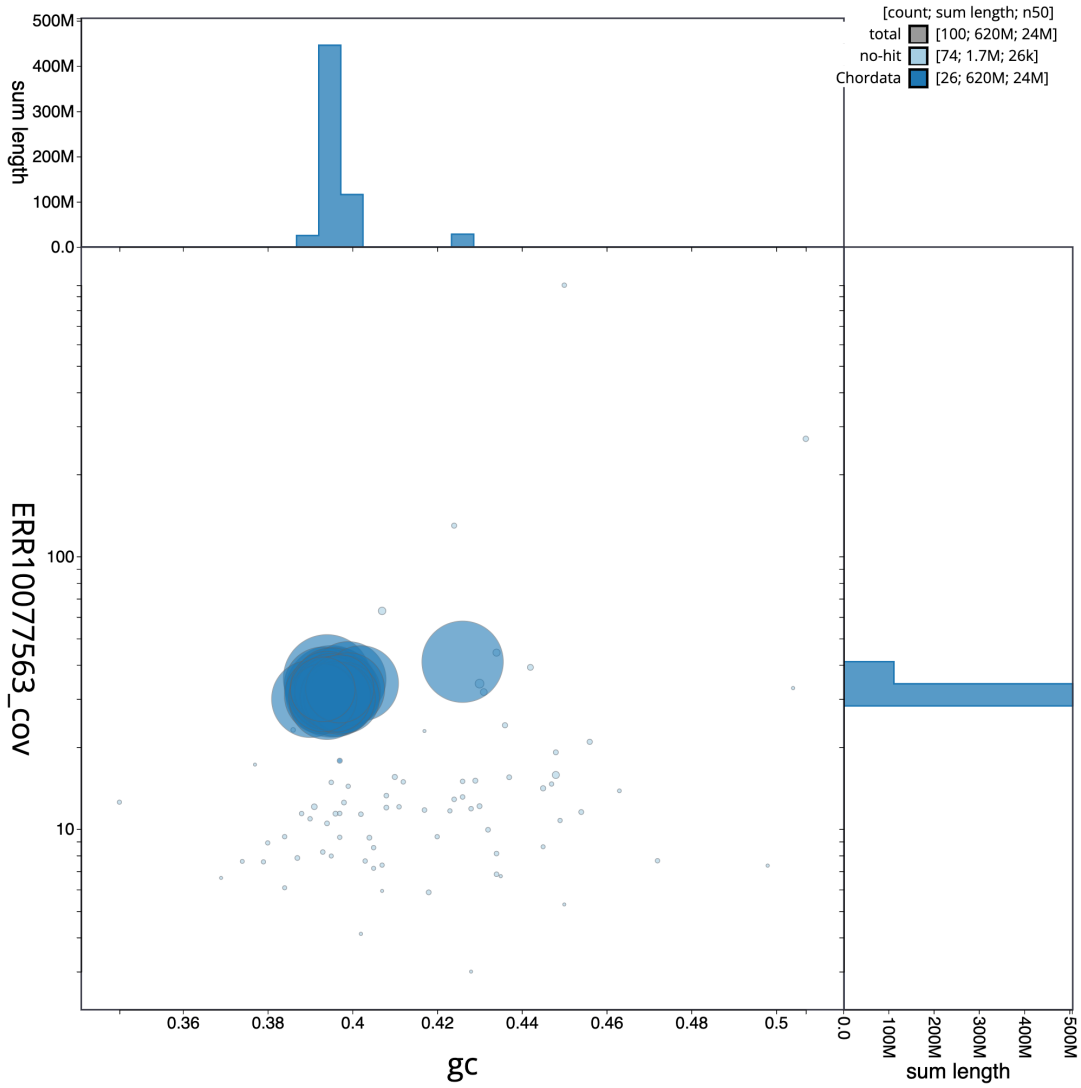

Figure 3. Genome assembly of *Barbatula barbatula*, fBarBar1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Barbatula%20barbatula/dataset/CAMQQH01/blob.

was transported alive to the University of Hull and left to recover fully in an aquarium before any sampling commenced. The specimen was euthanised in a lethal dose of MS-222 and tissue dissection was carried out by Bernd Hänfling within 30 minutes of euthanasia, and the tissues were immediately shock-frozen in liquid nitrogen.

High molecular weight (HMW) DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI), following a sequence of core procedures: sample preparation; sample homogenisation; HMW DNA extraction; DNA fragmentation; and DNA clean-up. Heart tissue of the fBarBar1sample was weighed and dissected on dry ice (as per the protocol https://dx.doi.org/10.17504/protocols.io.x54v9prmqg3e/v1). The tissue was homogenised using a Nippi Powermasher fitted with a BioMasher pestle, following the protocol at https://dx.doi.org/10.17504/protocols.io.5qpvo3r19v4o/v1. DNA was extracted by means of the HMW DNA Extraction: Automated MagAttract protocol (https://dx.doi.org/10.17504/protocols.io.kxygx3y4dg8j/v1). HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30, following the HMW DNA Fragmentation: Diagenode Megaruptor®3 for PacBio HiFi protocol (https://dx.doi.org/10.17504/protocols.io.8epv5x2zjg1b/v1). Sheared DNA was purified using Manual solid-phase reversible immobilisation (SPRI) (protocol at https://dx.doi.org/10.17504/protocols.io.kxygx3y1dg8j/v1). In brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter

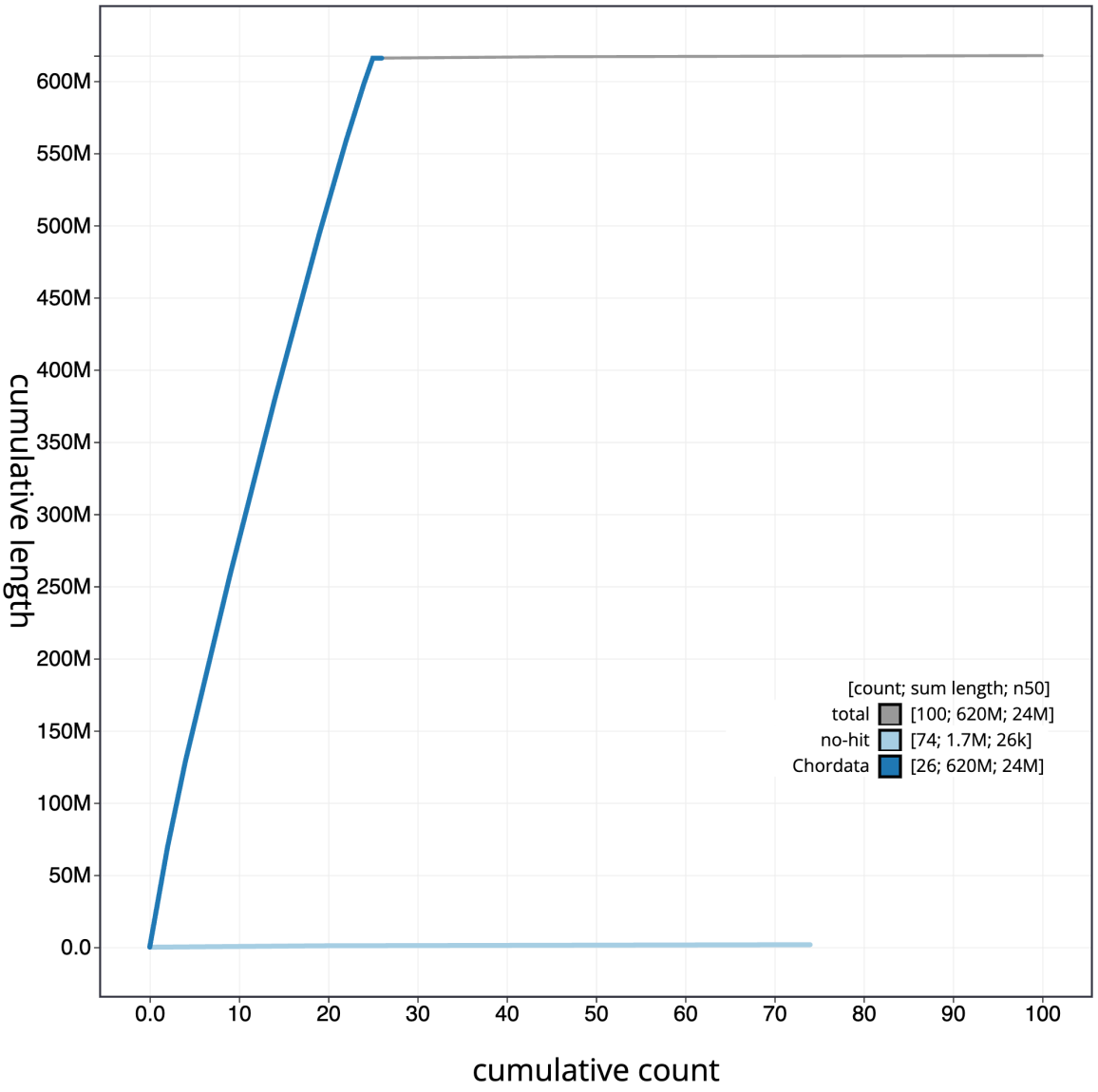

Figure 4. Genome assembly of *Barbatula barbatula*, fBarBar1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Barbatula%20barbatula/dataset/CAMQQH01/cumulative.

fragments and concentrate the DNA. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from spleen tissue of fBarBar1 using the Automated MagMax™ *mir*Vana protocol (https://dx.doi.org/10.17504/protocols.io.6qpvr36n3vmk/v1). The RNA concentration was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Protocols employed by the Tree of Life laboratory are publicly available on protocols.io: https://dx.doi.org/10.17504/protocols.io.8epv5xxy6g1b/v1.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed

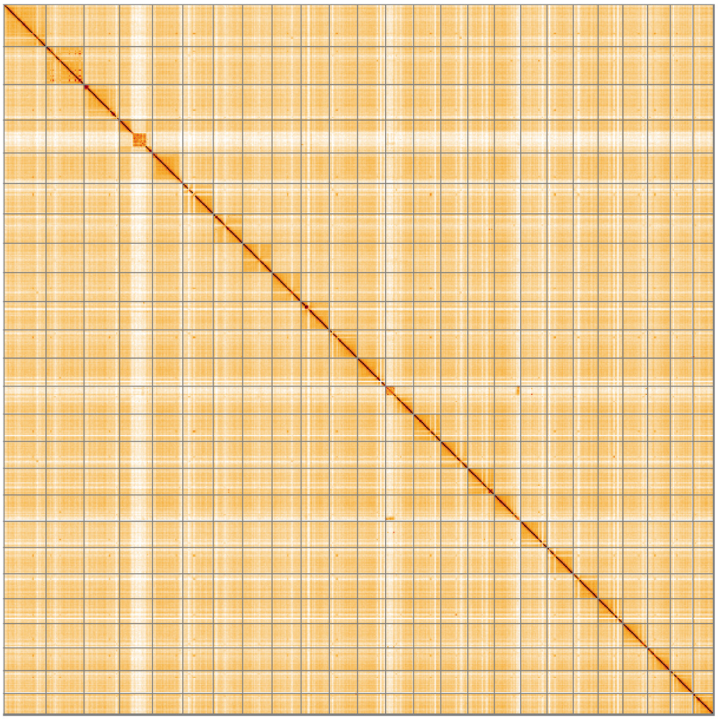

Figure 5. Genome assembly of *Barbatula barbatula*, fBarBar1.1: Hi-C contact map of the fBarBar1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=cvQOJvgKRwiPDDvhlbadUA.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Barbatula barbatula*, fBarBar1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX344768.1	1	36.42	39.5
OX344769.1	2	33.03	39.5
OX344770.1	3	30.81	39.5
OX344771.1	4	28.56	42.5
OX344772.1	5	26.42	39.5
OX344773.1	6	26.4	39.5
OX344774.1	7	25.59	39.5
OX344775.1	8	25.44	39.0
OX344776.1	9	25.21	39.5
OX344777.1	10	24.51	40.0
OX344778.1	11	24.5	39.5
OX344779.1	12	24.42	39.5
OX344780.1	13	24.24	40.0
OX344781.1	14	23.69	39.5
OX344782.1	15	23.35	39.5
OX344783.1	16	23.13	39.5
OX344784.1	17	22.89	40.0
OX344785.1	18	22.83	39.5
OX344786.1	19	22.75	40.0
OX344787.1	20	21.64	40.0
OX344788.1	21	21.57	39.5
OX344789.1	22	21.27	39.5
OX344791.1	24	19.77	39.5
OX344790.1	23	19.74	39.5
OX344792.1	25	17.71	39.5
OX344793.1	MT	0.02	45.5

according to the manufacturers' instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from muscle tissue of fBarBar1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) or MITOS (Bernt *et al.*, 2013) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines "sanger-tol/readmapping" (Surana *et al.*, 2023a) and "sanger-tol/genomenote" (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

Table 3 contains a list of relevant software tool versions and sources.

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 here. 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)

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

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: *Barbatula barbatula* (stone loach). Accession number PRJEB55341; https://identifiers.org/ena.embl/PRJEB55341 (Wellcome Sanger Institute, 2022). The genome sequence is released openly for reuse. The *Barbatula barbatula* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. 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.

Author information

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: https://doi.org/10.5281/zenodo.4783585.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: https://doi.org/10.5281/zenodo.4790455.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

Acknowledgements

Rob Donnelly (University of Hull) has provided assistance with tissue dissections and Alan Smith (University of Hull) during husbandry.

References

Abdennur N, Mirny LA: **Cooler: Scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Barluenga M, Meyer A: **Old fish in a young lake: stone loach (Pisces: *Barbatula barbatula*) populations in Lake Constance are genetically isolated by distance.** *Mol Ecol.* 2005; **14**(4): 1229–1239.
PubMed Abstract | Publisher Full Text

Bernt M, Donath A, Jühling F, *et al.*: **MITOS: Improved *de novo* metazoan mitochondrial genome annotation.** *Mol Phylogenet Evol.* 2013; **69**(2): 313–319.
PubMed Abstract | Publisher Full Text

Bubb DH, O'Malley OJ, Gooderham AC, *et al.*: **Relative impacts of native and non-native crayfish on shelter use by an indigenous benthic fish.** *Aquat Conserv.* 2009; **19**(4): 448–455.
Publisher 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

Copp GH, Vilizzi L: **Spatial and ontogenetic variability in the microhabitat use of stream-dwelling spined loach (*Cobitis taenia*) and stone loach (*Barbatula barbatula*).** *J Appl Ichthyol.* 2004; **20**(6): 440–451.
Publisher Full Text

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | Publisher Full Text

Freyhof J: **Barbatula barbatula.** IUCN Red List of Threatened Species 2011. 2011; [Accessed 12 October 2023].
Reference Source

Freyhof J, Brooks E: **European Red List of Freshwater Fishes.** Luxembourg: European Commission, 2011.
Reference Source

Froese R, Pauly D: **FishBase.** 2023.
Reference Source

Galib SM, Findlay JS, Lucas MC: **Strong impacts of signal crayfish invasion on upland stream fish and invertebrate communities.** *Freshwater Biology.* 2021; **66**(2): 223–240.
Publisher 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

Guan RZ, Wiles PR: **Ecological impacts of introduced crayfish on benthic fishes in a British lowland river.** *Conservation Biology.* 1997; **11**(3): 641–647.
Reference Source

Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022].
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* Oxford University Press, 2021; **10**(1): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Hyslop EJ: **The feeding-habits of 0+ stone loach, *Noemacheilus barbatulus* (L.), and bullhead, Cottus gobio L.** *J Fish Biol.* 1982; **21**(2): 187–196.
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

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

Kottelat M, Freyhof J: **Handbook of European Freshwater Fishes.** Berlin: Publications Kottelat, 2007b.
Reference Source

Kováč V, Copp GH, Francis MP: **Morphometry of the stone loach, *Barbatula barbatula*: do mensural characters reflect the species' life history thresholds?** *Environ Biol Fishes.* 1999; **56**: 105–115.
Publisher Full Text

MacKenzie AR, Greenberg L: **The influence of instream cover and predation**

risk on microhabitat selection of stone loach *Barbatula barbatula* (L.). *Ecol Freshw Fish.* 1988; **7**: 87–94.

Maitland P: **Keys to the Freshwater Fish of Great Britain and Ireland with Notes on their Distribution and Ecology.** *Scientific Publication No. 62.* Ambleside: Freshwater Biological Association, 2004.

Maitland PS: **The feeding relationships of salmon, trout, minnows, stone loach and three-spined sticklebacks in the River Endrick, Scotland.** *J Anim Ecol.* 1965; **34**(1): 109–133.
Publisher Full Text

Mann RHK: **The populations, growth and production of fish in four small streams in southern England.** *J Anim Ecol.* 1971; **40**(1): 155–190.
Publisher Full Text

Mann RHK, Mills CA, Crisp DT: **Geographical variation in the life-history tactics of some species of freshwater fish.** In: Potts, G. W. and Wootton, R. J. (eds.) *Fish Reproduction: Strategies and Tactics.* London: Academic Press: 1984; 171–186.

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

Mills CA, Welton JS, Rendle EL: **The age, growth and reproduction of the stone loach *Noemacheilus barbatulus* (L.) in a Dorset chalk stream.** *Freshwater Biology.* 1983; **13**(3): 283–292.
Publisher Full Text

Mills CA, Eloranta A: **Reproductive Strategies in the Stone Loach *Noemacheilus barbatulus*.** *Oikos.* 1985; **44**(2): 341–349.
Publisher Full Text

Nunn AD, Copp GH, Vilizzi L, *et al.*: **Seasonal and diel patterns in the migrations of fishes between a river and a floodplain tributary.** *Ecol Freshw Fish.* 2010; **19**(1): 153–162.
Publisher Full Text

Nunn AD, Harvey JP, Cowx IG: **Benefits to 0+ fishes of connecting man-made waterbodies to the lower River Trent, England.** *River Res Appl.* 2007; **23**(4): 361–376.
Publisher Full Text

Prenda J, Armitage PD, Grayston A: **Habitat use by the fish assemblages of two chalk streams.** *J Fish Biol.* 1997; **51**(1): 64–79.
PubMed Abstract | Publisher 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

Šedivá A, Janko K, Šlechtová V, *et al.*: **Around or across the Carpathians: colonization model of the Danube basin inferred from genetic diversification of stone loach (*Barbatula barbatula*) populations.** *Mol Ecol.* 2008; **17**(5): 1277–1292.
PubMed Abstract | Publisher Full Text

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**(19): 3210–3212.
PubMed Abstract | Publisher Full Text

Skryabin AG: **The biology of stone loach *Barbatula barbatulus* in the Rivers Goloustnaya and Olkha, East Siberia.** *J Fish Biol.* 1993; **42**(3): 361–374.
Publisher Full Text

Smyly WJP: **On the biology of the stone-loach *Nemacheilus barbatula* (L.).** *J Anim Ecol.* 1955; **24**(1): 167–186.
Publisher Full Text

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a; [Accessed 21 July 2023].
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b; [Accessed 21 July 2023].
Publisher Full Text

Tudorache C, Viaene P, Blust R, *et al.*: **A comparison of swimming capacity and energy use in seven European freshwater fish species.** *Ecol Freshw Fish.* 2008; **17**(2): 284–291.
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

Vinyoles D, De Sostoa A, Franch C, *et al.*: **Life-history traits of the stone loach *Barbatula barbatula*.** *J Fish Biol.* 2010; **77**(1): 20–32.
PubMed Abstract | Publisher Full Text

Welton JS, Mills CA, Rendle EL: **Food and habitat partitioning in two small benthic fishes, *Noemacheilus barbatulus* (L.) and *Cottus gobio* L.** *Arch Hydrobiol.* 1983; **97**(4): 434–454.
Reference Source

Wheeler A: **Freshwater fishes of Britain and Europe.** London: Rainbow Books, 1992.

Wellcome Sanger Institute: **The genome sequence of the stone loach, *Barbatula barbatula* (Linnaeus, 1758). European Nucleotide Archive.** [dataset], accession number PRJEB55341, 2022.

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

Zweimüller I: **Microhabitat use by two small benthic stream fish in a 2nd order stream.** *Hydrobiologia.* 1995; **303**: 125–137.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 02 January 2024

https://doi.org/10.21956/wellcomeopenres.22474.r70459

✔ **Guoqing Lu**
Department of Biology, University of Nebraska Omaha, Omaha, Nebraska, USA

The paper presents the first draft genome of an individual female stone loach (*Barbatula barbatula*), a fish species belonging to Cypriniformes. The assembly spans 617.6 Mb and is organized into 25 pseudochromosomes. It would be more helpful if the whole genome were annotated and a comparative genomic analysis were conducted. Nevertheless, this study provides the first reference genome for the stone loach, a valuable resource for future research on fish biodiversity, ecology, and evolution.

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, Bioinformatics, Fish Biology

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.