DATA NOTE

The genome sequence of the gudgeon, *Gobio gobio* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]

Andy D. Nunn[1], Paolo Moccetti[1], Bernd Hänfling[2],
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]School of Natural Sciences, University of Hull, Hull, England, UK
[2]Institute for Biodiversity and Freshwater Conservation, University of Highlands and Islands, Inverness, Scotland, UK

v1 **First published:** 26 Jul 2024, **9**:409
https://doi.org/10.12688/wellcomeopenres.22742.1
Latest published: 26 Jul 2024, **9**:409
https://doi.org/10.12688/wellcomeopenres.22742.1

Abstract

We present a genome assembly from an individual female *Gobio gobio* (the gudgeon; Chordata; Actinopteri; Cypriniformes; Gobionidae). The genome sequence spans 1,460.70 megabases. Most of the assembly is scaffolded into 25 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 16.61 kilobases in length.

Keywords

Gobio gobio, gudgeon, genome sequence, chromosomal, Cypriniformes

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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: Investigation, Resources, Writing – Original Draft Preparation; **Moccetti P**: Resources; **Hänfling B**: Investigation, 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 [206194, https://doi.org/10.35802/206194] 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: Nunn AD, Moccetti P, Hänfling B *et al.* **The genome sequence of the gudgeon, *Gobio gobio* (Linnaeus, 1758) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:409
https://doi.org/10.12688/wellcomeopenres.22742.1

First published: 26 Jul 2024, **9**:409 https://doi.org/10.12688/wellcomeopenres.22742.1

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; Gobionidae; Gobioninae; *Gobio; Gobio gobio* (Linnaeus, 1758) (NCBI:txid27704).

Background

Gudgeon *Gobio gobio* (L.) occur across much of central Europe and Russia to the west of the Ural Mountains (Kottelat & Freyhof, 2007). The species is a freshwater resident, has a single barbel on either side of the mouth and a maximum body length of ~20 cm (Maitland, 2004). Life span and age- and size-at-maturity are typically 5–8 years, 7–10 cm at 2–3 years respectively, but are strongly influenced by environmental conditions (Lobon-Cervia *et al.*, 1991; Maitland, 2004). The taxonomy of the genus is complex (Mendel *et al.*, 2008; Takács *et al.*, 2014) and the species' geographical distribution incompletely understood, with several taxa probably included under a single name in some locations (Kottelat & Freyhof, 2007).

Gudgeon invariably inhabit glides in the middle and lower reaches of rivers (Mann, 1980; Prenda *et al.*, 1997), but can also occur in connected still waters (Nunn *et al.*, 2007a). Most individuals are comparatively sedentary, although small-scale movements and shifts to deeper, faster-flowing water with increasing fish size have been documented (Nunn *et al.*, 2010; Stott, 1967; Watkins *et al.*, 1997). The species is usually iteroparous (but see Lobón-Cerviá *et al.*, 1991), with batches of eggs deposited on sand, or sometimes gravel or aquatic vegetation, in flowing water over the spring and summer (Mann, 1980; Nunn *et al.*, 2007c). Gudgeon consume mainly small crustaceans, insect larvae, molluscs and algae, irrespective of season and developmental period (Hartley, 1948; Nunn *et al.*, 2007b).

According to the International Union for the Conservation of Nature (IUCN) Red List of Threatened Species, gudgeon is classified as "Least Concern" in terms of extinction risk and no major threats have been documented (Freyhof, 2011; Freyhof & Brooks, 2011; Nunn *et al.*, 2023). Nonetheless, some populations are undoubtedly exposed to and compromised by a range of common and widespread pressures, including pollution and habitat degradation and fragmentation (Bervoets & Blust, 2003; Blanchet *et al.*, 2010; Faller *et al.*, 2003; Knaepkens *et al.*, 2007).

The genome of the gudgeon, *Gobio gobio*, was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland. Here we present a chromosomally complete genome sequence for *Gobio gobio*, based on one female specimen from the River Wharfe, UK.

Genome sequence report

The genome of an adult *Gobio gobio* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 31.40 Gb (gigabases) from

Figure 1. Photograph of the *Gobio gobio* (fGobGob1) specimen used for genome sequencing.

2.22 million reads, providing approximately 34-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 161.52 Gbp from 1,069.67 million reads, yielding an approximate coverage of 111-fold. Specimen and sequencing information is summarised in Table 1.

Manual assembly curation corrected 72 missing joins or mis-joins and 39 haplotypic duplications, reducing the assembly length by 1.88% and the scaffold number by 21.69%, and increasing the scaffold N50 by 6.39%. The final assembly has a total length of 1,460.70 Mb in 230 sequence scaffolds with a scaffold N50 of 58.1 Mb (Table 2). The snail plot in Figure 2 provides a summary of the assembly statistics, while Figure 3 shows the distribution of assembly scaffolds based on base coverage across chromosomes. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (98.82%) 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 3). 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 57.3 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.3.2 completeness of 96.9% (single = 95.2%, duplicated = 1.7%), using the actinopterygii_odb10 reference set (n = 3,640).

Metadata for specimens, BOLD barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://tolqc.cog.sanger.ac.uk/darwin/fish/Gobio_gobio/.

Methods

Sample acquisition and nucleic acid extraction

A female adult specimen of *G. gobio* (specimen ID SAN0000704, ToLID fGobGob1) was collected from the River Wharfe, UK (latitude 53.91, longitude –1.61) on 2020-09-09.

Table 1. **Specimen and sequencing data for *Gobio gobio*.**

Project information			
Study title	Gobio gobio (gudgeon)		
Umbrella BioProject	PRJEB59786		
Species	*Gobio gobio*		
BioSample	SAMEA11296538		
NCBI taxonomy ID	27704		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	fGobGob1	SAMEA11296590	spleen
Hi-C sequencing	fGobGob1	SAMEA11296597	gill
RNA sequencing	fGobGob1	SAMEA11296588	heart
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR10890732	1.07e+09	161.52
PacBio Sequel IIe	ERR10879931	1.76e+06	19.79
PacBio Sequel IIe	ERR10879930	2.22e+06	31.4
RNA Illumina NovaSeq 6000	ERR10890733	6.96e+07	10.51

Table 2. **Genome assembly data for *Gobio gobio*, fGobGob1.1.**

Genome assembly		
Assembly name	fGobGob1.1	
Assembly accession	GCA_949357685.1	
Accession of alternate haplotype	*GCA_949357705.1*	
Span (Mb)	1,460.70	
Number of contigs	1,277	
Contig N50 length (Mb)	2.6	
Number of scaffolds	230	
Scaffold N50 length (Mb)	58.1	
Longest scaffold (Mb)	87.33	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	57.3	*≥ 50*
k-mer completeness	99.99%	*≥ 95%*
BUSCO**	C:96.9%[S:95.2%,D:1.7%], F:1.2%,M:1.9%,n:3,640	*C ≥ 95%*

Assembly metrics*		***Benchmark***
Percentage of assembly mapped to chromosomes	98.82%	*≥ 95%*
Sex chromosomes	Not identified	*localised homologous pairs*
Organelles	Mitochondrial genome: 16.61 kb	*complete single alleles*

* 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_odb10 BUSCO set using version 5.3.2 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/Gobio%20gobio/dataset/CASHTD01.1/busco.

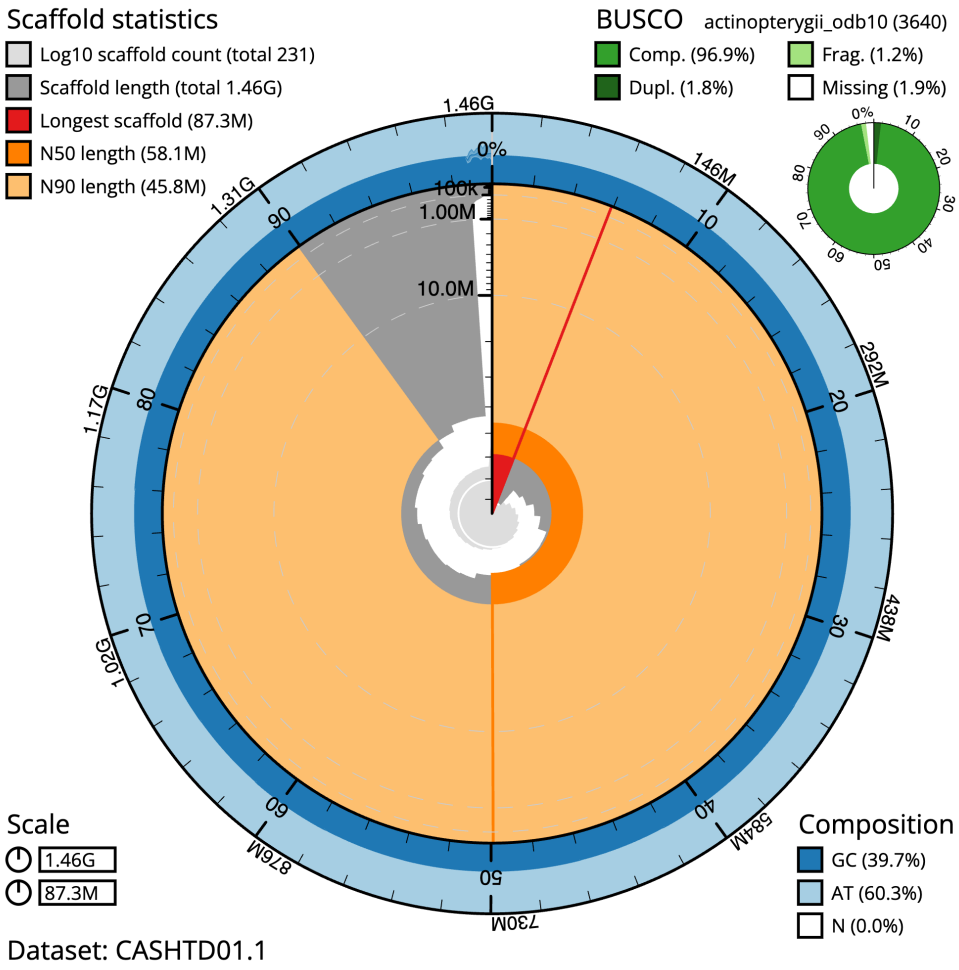

Figure 2. Genome assembly of *Gobio gobio*, fGobGob1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,460,679,231 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (87,331,821 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (58,069,932 and 45,801,292 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/Gobio%20gobio/dataset/CASHTD01.1/snail.

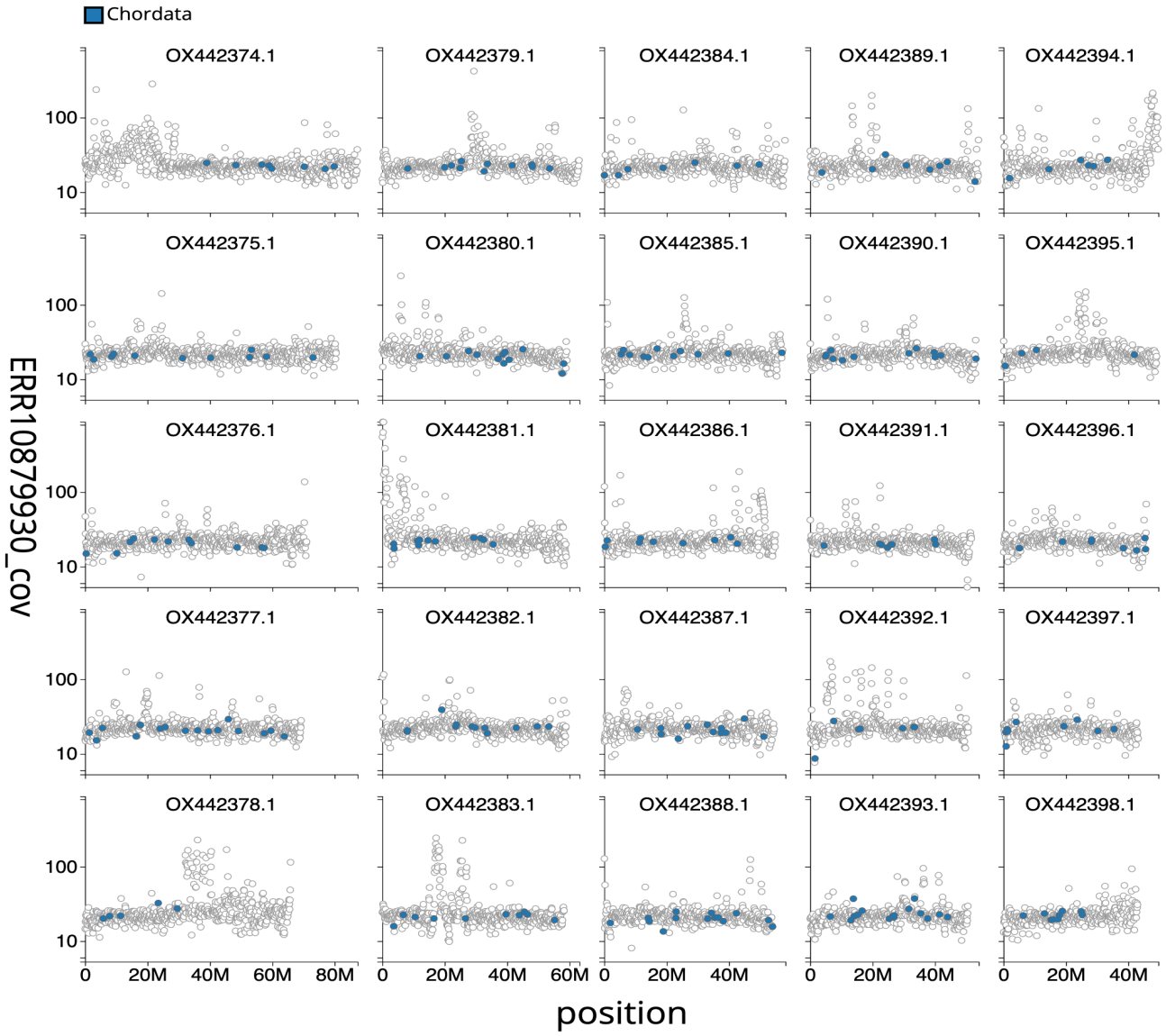

Figure 3. Genome assembly of *Gobio gobio*, fGobGob1.1: Distribution plot of base coverage in ERR10879930 against position for sequences in assembly CASHTD01.1. Windows of 100kb are coloured by phylum. The assembly has been filtered to exclude sequences with length < 2,550,000. An interactive version of this figure is available here.

The specimen was collected by Andy Nunn and Paolo Moccetti using electro-fishing, and identified by Andy Nunn and Bernd Hänfling. The specimen was transported alive to the University of Hull and left to recover fully in an aquarium before any sampling commenced. The specimen was euthanized 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.

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the fGobGob1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the spleen was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Automated MagAttract v1 protocol (Sheerin *et al.*, 2023). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30 (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X

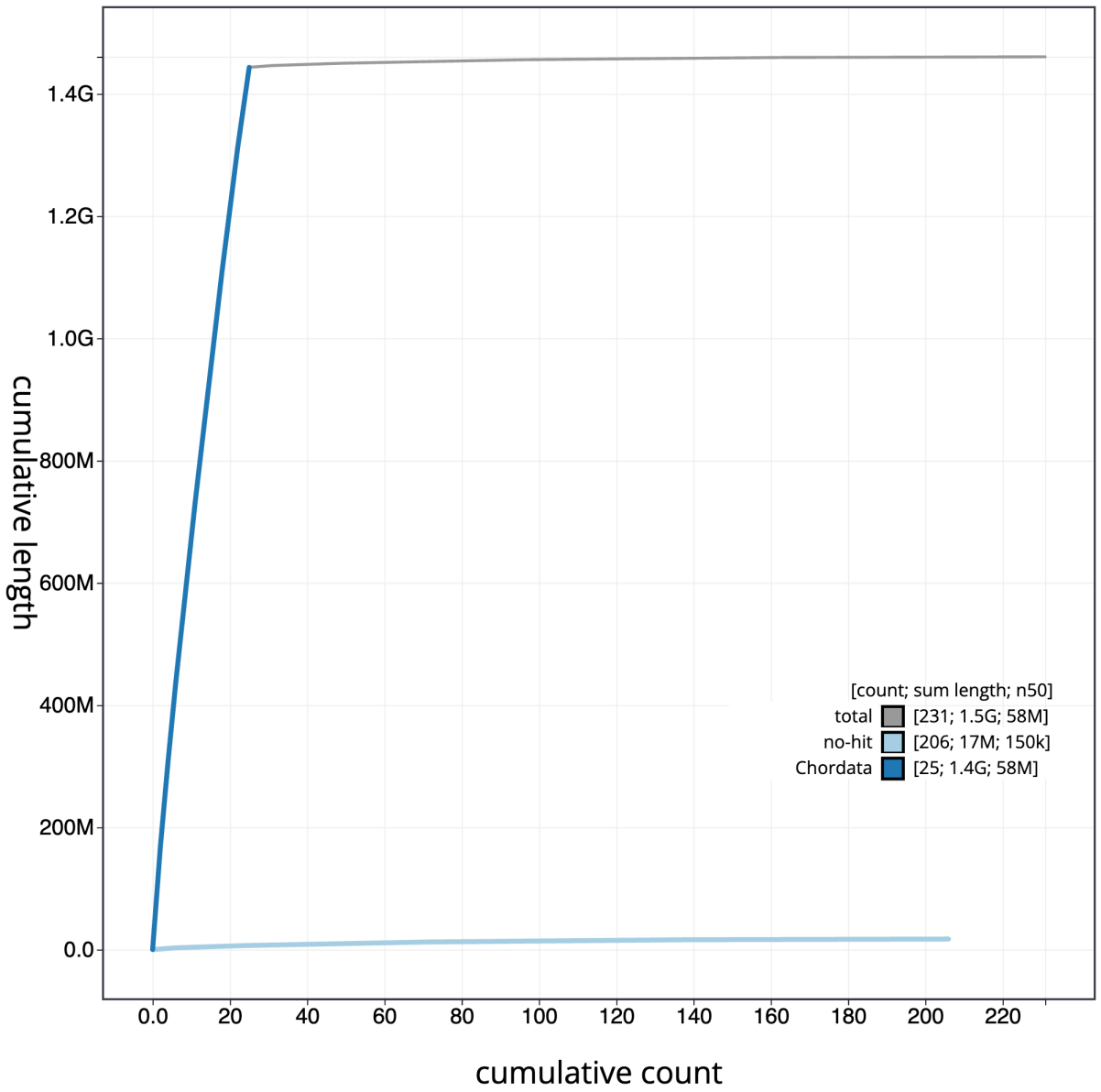

Figure 4. Genome assembly of *Gobio gobio* fGobGob1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all sequences. Coloured lines show cumulative lengths of sequences assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Gobio%20gobio/dataset/CASHTD01.1/cumulative.

ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. 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.

RNA was extracted from heart tissue of fGobGob1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). 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.

Protocols developed by the WSI Tree of Life laboratory are publicly available on protocols.io (Denton *et al.*, 2023b).

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers'

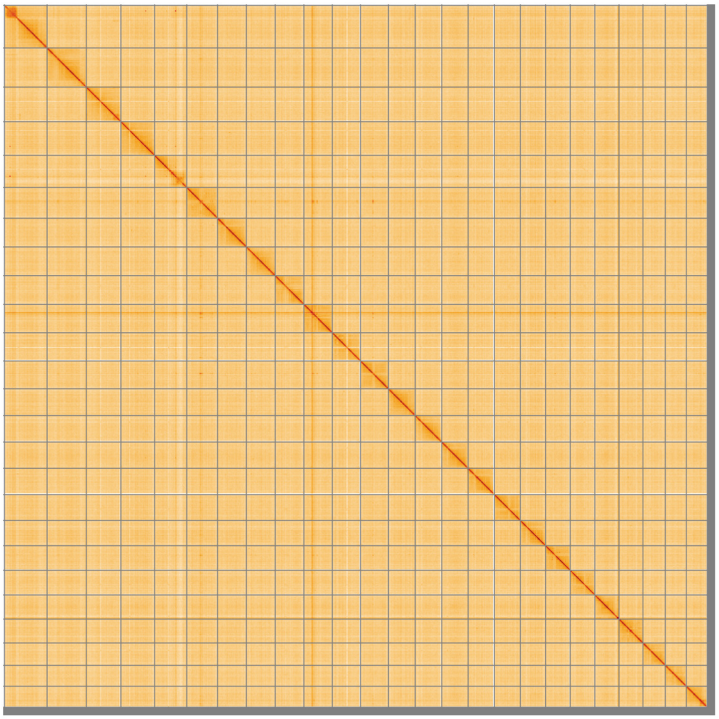

Figure 5. Genome assembly of *Gobio gobio* fGobGob1.1: Hi-C contact map of the fGobGob1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=YsWLoMZ1TAiTKAKnA0qYxw.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Gobio gobio*, fGobGob1.

INSDC accession	Name	Length (Mb)	GC%
OX442374.1	1	87.33	40.0
OX442375.1	2	80.6	40.0
OX442376.1	3	71.07	39.5
OX442377.1	4	69.34	39.5
OX442378.1	5	65.97	40.0
OX442379.1	6	63.19	40.0
OX442380.1	7	59.21	39.5
OX442381.1	8	59.06	40.0
OX442382.1	9	59.04	40.0
OX442383.1	10	58.33	39.5
OX442384.1	11	58.07	39.5
OX442385.1	12	57.3	39.5
OX442386.1	13	54.98	39.5
OX442387.1	14	54.58	40.0
OX442388.1	15	54.08	39.5
OX442389.1	16	54.03	39.5
OX442390.1	17	53.27	39.5
OX442391.1	18	51.76	39.0
OX442392.1	19	50.72	40.0
OX442393.1	20	50.67	39.5
OX442394.1	21	49.65	40.0
OX442395.1	22	49.16	40.0
OX442396.1	23	45.8	39.5
OX442397.1	24	43.23	40.0
OX442398.1	25	43.01	40.0
OX442399.1	MT	0.02	44.0

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 IIe (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from gill tissue of fGobGob1 using the Arima-HiC v2 kit. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly. Original assembly of HiFi reads is performed using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed with purge_dups (Guan *et al.*, 2020). Hi-C reads are further mapped with bwa-mem2 (Vasimuddin *et al.*, 2019) to the primary contigs, which are further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option. Scaffolded assemblies are 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 (article in preparation). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were corrected, and duplicate sequences were tagged and removed. The entire process is documented at https://gitlab.com/wtsi-grit/rapid-curation (article in preparation).

Evaluation of the final assembly. 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.

The readmapping pipelines were developed using the nf-core tooling (Ewels *et al.*, 2020), use MultiQC (Ewels *et al.*, 2016), and make extensive use of the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), and the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 4 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

Table 4. 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	1.2a	https://github.com/c-zhou/yahs

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.

Data availability

European Nucleotide Archive: *Gobio gobio* (gudgeon). Accession number PRJEB59786; https://identifiers.org/ena.embl/PRJEB59786 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Gobio gobio* 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 and Table 2.

Author information

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

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) 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

Bervoets L, Blust R: **Metal concentrations in water, sediment and gudgeon (*Gobio gobio*) from a pollution gradient: relationship with fish condition factor.** *Environ Pollut.* 2003; **126**(1): 9–19.
PubMed Abstract | Publisher Full Text

Blanchet S, Rey O, Etienne R, *et al.*: **Species-specific responses to landscape fragmentation: implications for management strategies.** *Evol Appl.* 2010; **3**(3): 291–304.
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

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.
PubMed Abstract | Publisher Full Text | Free Full Text

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
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

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
PubMed Abstract | Publisher Full Text | Free Full Text

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.
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

Faller P, Kobler B, Peter A, *et al.*: **Stress status of gudgeon (*Gobio gobio*) from rivers in Switzerland with and without input of sewage treatment plant effluent.** *Environ Toxicol Chem.* 2003; **22**(9): 2063–72.
PubMed Abstract | 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

Freyhof J: ***Gobio gobio*, IUCN Red List of Threatened Species 2011.** 2011; [Accessed 16 June 2024].
Publisher Full Text

Freyhof J, Brooks E: **European red list of freshwater fishes.** Luxembourg, 2011.
Reference Source

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

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

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

Hartley PHT: **Food and feeding relationships in a community of fresh-water fishes.** *J Anim Ecol.* 1948; **17**(1): 1–14.
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

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
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

Knaepkens G, Maerten E, Eens M: **Performance of a pool-and-weir fish pass for small bottom-dwelling freshwater fish species in a regulated lowland river.** *Anim Biol.* 2007; **57**(4): 423–432.
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

Lobon-Cervia J, Montañges C, De Sostoa A: **Influence of environment upon the life history of gudgeon, *Gobio gobio* (L.): a recent and successful colonizer of the Iberian Peninsula.** *J Fish Biol.* 1991; **39**(3): 285–300.
Publisher Full Text

Maitland PS: **Keys to the freshwater fish of Britain and Ireland, with notes on their distribution and ecology.** Sutcliffe, D. W. (ed.). Ambleside, England: Freshwater Biological Association, 2004.
Reference Source

Mann RHK: **The growth and reproductive strategy of the gudgeon, *Gobio gobio* (L.), in two hard-water rivers in southern England.** *J Fish Biol.* 1980; **17**(2): 163–176.
Publisher Full Text

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

Mendel J, Lusk S, Vasil'eva ED, *et al.*: **Molecular phylogeny of the genus *Gobio* Cuvier, 1816 (Teleostei: Cyprinidae) and its contribution to taxonomy.** *Mol Phylogenet Evol.* 2008; **47**(3): 1061–75.
PubMed Abstract | Publisher Full Text

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

Nunn AD, Ainsworth RF, Walton S, *et al.*: **Extinction risks and threats facing the freshwater fishes of Britain.** *Aquat Conserv.* 2023; **33**(12): 1460–1476.
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.* 2007a; **23**(4): 361–376.
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

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.* 2007c; **70**(4): 1254–1267.
Publisher Full Text

Prenda J, Armitage P, 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

Sheerin E, Sampaio F, Oatley G, *et al.*: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.1.** *protocols.io.* 2023.
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

Stott B: **The Movements and population densities of roach (*Rutilus rutilus* (L.)) and gudgeon (*Gobio gobio* (L.)) in the River Mole.** *J Anim Ecol.* 1967; **36**(2): 407–423.
Publisher Full Text

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023.
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.
Publisher Full Text

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

Takács P, Bihari P, Erős T, *et al.*: **Genetic heterogeneity reveals on-going speciation and cryptic taxonomic diversity of stream-dwelling gudgeons (Teleostei, Cyprinidae) in the middle Danubian hydrosystem (Hungary).** *PLoS One.* 2014; **9**(5): e97278.
PubMed Abstract | Publisher Full Text | Free Full Text

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023.
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

Watkins MS, Doherty S, Copp GH: **Microhabitat use by 0+ and older fishes in a small English chalk stream.** *J Fish Biol.* 1997; **50**(5): 1010–1024.
Publisher Full Text

Wellcome Sanger Institute: **The genome sequence of the gudgeon, *Gobio gobio* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB59786, 2023.

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