Check for updates

DATA NOTE

The genome sequence of the European conger eel, *Conger conger* (Linnaeus, 1758) [version 1; peer review: 3 approved]

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

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

Abstract

We present a genome assembly from an individual *Conger conger* (the European conger eel; Chordata; Actinopteri; Anguilliformes; Congridae). The genome sequence spans 1,136.40 megabases. Most of the assembly is scaffolded into 19 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 18.86 kilobases in length.

Keywords

Conger conger, European conger, genome sequence, chromosomal, Anguilliformes

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 1 18 Sep 2024	✓ view	✓ view	✓ view

1. **Chong Han**, Guangzhou University, Guangzhou, China

2. **Khairul Syahputra**, National Research and Innovation Agency (BRIN), Cibinong, Indonesia

3. **Kazuharu Nomura**, Fisheries Technology Institute, Japan Fisheries Research and Education Agency, Mie, Japan

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: Adkins P: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Brittain R**: Investigation, Resources; **Harley J**: Investigation, Resources; **Modepali V**: Writing – Original Draft Preparation, Writing – Review & Editing;

How to cite this article: Adkins P, Brittain R, Harley J *et al.* **The genome sequence of the European conger eel,** *Conger conger* **(Linnaeus, 1758) [version 1; peer review: 3 approved]** Wellcome Open Research 2024, **9**:532 https://doi.org/10.12688/wellcomeopenres.23052.1

First published: 18 Sep 2024, **9**:532 https://doi.org/10.12688/wellcomeopenres.23052.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Actinopterygii; Actinopteri; Neopterygii; Teleostei; Elopocephalai; Elopocephala; Elopomorpha; Anguilliformes; Congridae; Congrinae; Conger; *Conger conger* (Linnaeus, 1758) (NCBI:txid82655).

Background

The European Conger Eel *Conger conger* (Linnaeus, 1758) is the largest eel species found in Europe, distributed across the North-East Atlantic, Mediterranean, and western Black Sea (Whitehead, 1985). Conger eels are strictly marine benthic fishes that live on rocky and sandy bottoms to 500 depth, although they travel much deeper to spawn (Bauchot & Blache, 1980). Conger Eels are carnivores and mainly feed on bottom-living fishes, crustaceans, and cephalopods at night (Cau & Manconi, 1984; Levy *et al.*, 1988; Saldanha *et al.*, 1995). Conger eels are the largest of the family Congridae with records existing of specimens of over 2.7 m and weighing 65 kg (Wheeler, 1985), although such large fishes are uncommon (Fannon *et al.*, 1990). Male conger eels are reported to be smaller than females (Cau & Manconi, 1983). The European conger eel reaches sexual maturity at 5–15 years old and spawns terminally in deep waters during summer (Hayward & Ryland, 2017). Female Conger eels have a semelparous reproductive strategy, reproducing only once and dying after releasing several million eggs. Spawning grounds for conger eels have been identified in the Sardinian Channel, between Gibraltar and the Azores, and near the Azores archipelago at depths of up to 4000 m (Correia *et al.*, 2002; Correia *et al.*, 2009; Correia *et al.*, 2011; Correia *et al.*, 2012). Conger eel larvae are highly dispersive, and their development lasts for about 6–9 months before they metamorphose into juvenile eels (Correia *et al.*, 2011).

Conger eels are important commercial fishing species of the North-East Atlantic (Figueiredo *et al.*, 1996). Despite being a geographically widespread species and a valuable fisheries resource, relatively little is known about the reproductive biology, ecology, and migratory behaviour of Conger eels. The genome resource for the Conger eel adds valuable information to understanding their biology and population genetics, which could help in their conservation and management.

Genome sequence report

The genome of a juvenile *Conger conger* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 17.62 Gb (gigabases) from 1.46 million reads, providing approximately 36-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 253.94 Gbp from 1,681.70 million reads, yielding an approximate coverage of 223-fold. Specimen and sequencing information is summarised in Table 1.

Manual assembly curation corrected 23 missing joins or mis-joins, reducing scaffold number by 4.99%, and increasing
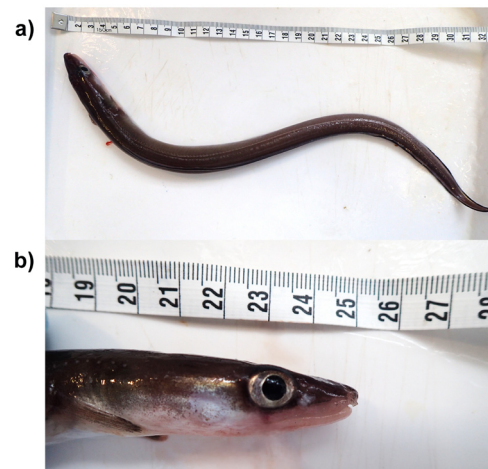
Figure 1. Photograph of the *Conger conger* (fConCon1) specimen used for genome sequencing.

the scaffold N50 by 37.08%. The final assembly has a total length of 1,136.40 Mb in 380 sequence scaffolds, with 1,009 gaps, and a scaffold N50 of 64.9 Mb (Table 2). The snail plot in Figure 2 provides a summary of the assembly 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 (97.58%) of the assembly sequence was assigned to 19 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 56.1 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.4.3 completeness of 94.9% (single = 87.4%, duplicated = 7.5%), 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://links.tol.sanger.ac.uk/species/82655.

Methods

Sample acquisition and barcoding

A juvenile *Conger conger* specimen (specimen ID MBA-210527-004A, ToLID fConCon1) was collected from Middle Ground, English Channel, UK (latitude 50.24, longitude –4.18) on 2021-05-27. The specimen was taken from its habitat Broken shell and muddy sand using an Agassiz trawl deployed from RV Sepia. The specimen was collected by Patrick Adkins and Joanna Harley (Marine Biological

**Table 1.** Specimen and sequencing data for *Conger conger*.

Project information	
Study title	Conger conger (European conger)
Umbrella BioProject	PRJEB65259
Species	*Conger conger*
BioSample	SAMEA12219431
NCBI taxonomy ID	82655

Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	fConCon1	SAMEA12219622	gill
Hi-C sequencing	fConCon1	SAMEA12219622	gill
RNA sequencing	fConCon1	SAMEA12219616	muscle

Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR11872601	1.68e+09	253.94
PacBio Sequel IIe	ERR11867230	2.08e+06	20.78
PacBio Sequel IIe	ERR11867231	1.46e+06	17.62
RNA Illumina NovaSeq 6000	ERR12245589	8.38e+07	12.65

Table 2. Genome assembly data for *Conger conger*, fConCon1.1.

Genome assembly	
Assembly name	fConCon1.1
Assembly accession	GCA_963514075.1
Accession of alternate haplotype	*GCA_963514125.1*
Span (Mb)	1,136.40
Number of contigs	1,390
Contig N50 length (Mb)	3.1
Number of scaffolds	380
Scaffold N50 length (Mb)	64.9
Longest scaffold (Mb)	96.78

Assembly metrics*		*Benchmark*
Consensus quality (QV)	56.1	*≥ 50*
k-mer completeness	99.99%	*≥ 95%*
BUSCO**	C:94.9%[S:87.4%,D:7.5%], F:1.8%,M:3.3%,n:3640	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	97.58%	*≥ 95%*
Sex chromosomes	Not identified	*localised homologous pairs*
Organelles	Mitochondrial genome: 18.86 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 5.4.3. 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/Conger_conger/dataset/GCA_963514075.1/busco.
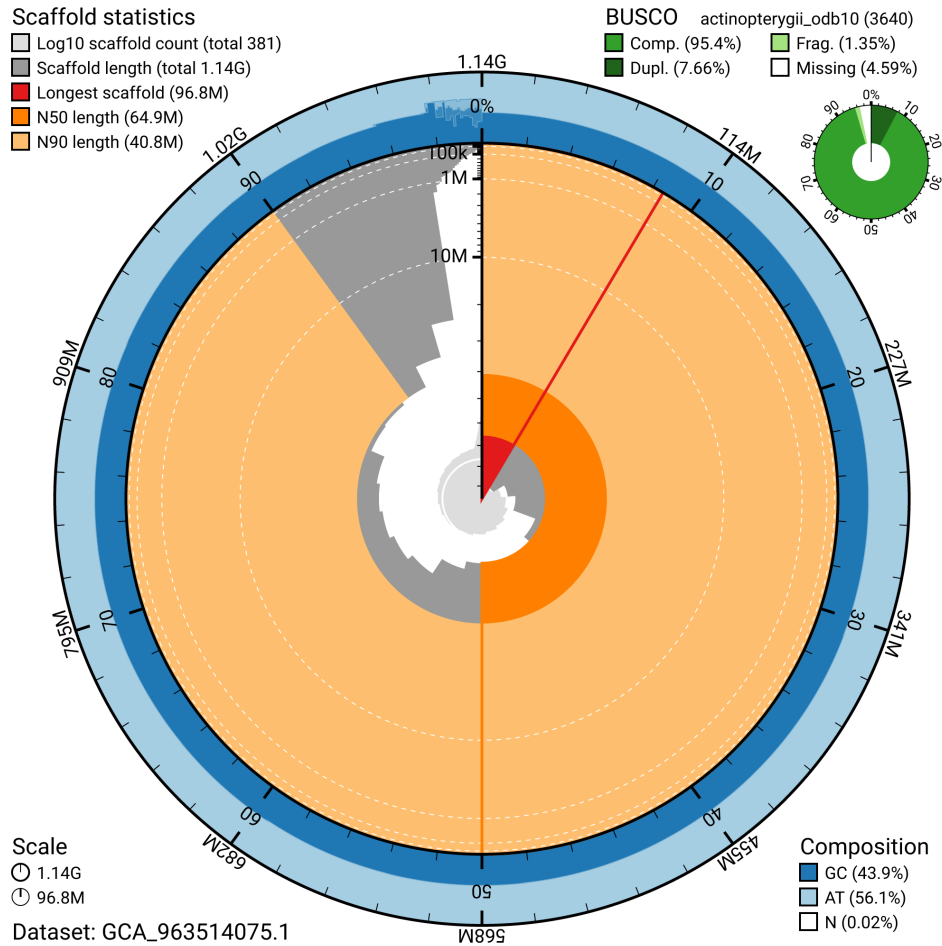
Figure 2. Genome assembly of *Conger conger*, fConCon1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,136,421,744 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (96,783,005 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (64,882,367 and 40,826,945 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/GCA_963514075.1/dataset/GCA_963514075.1/snail.

Association) and identified by Rachel Brittain (Marine Biological Association) based on gross morphology. The fish was first anesthetised and then overdosed using Aquased (2-phenoxyethanol). Destruction of the brain was used as a secondary method to ensure the animal was deceased before tissue sampling took place as in accordance with Schedule 1 methodology under the home office licence. Samples taken from the animal were preserved on dry ice.

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimens and stored in ethanol, while the remaining parts of the specimen were shipped on dry ice to the Wellcome

Sanger Institute (WSI). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding have been deposited on protocols.io (Beasley *et al.*, 2023).

Nucleic acid extraction
The workflow for high molecular weight (HMW) DNA extraction at the WSI Tree of Life Core Laboratory includes
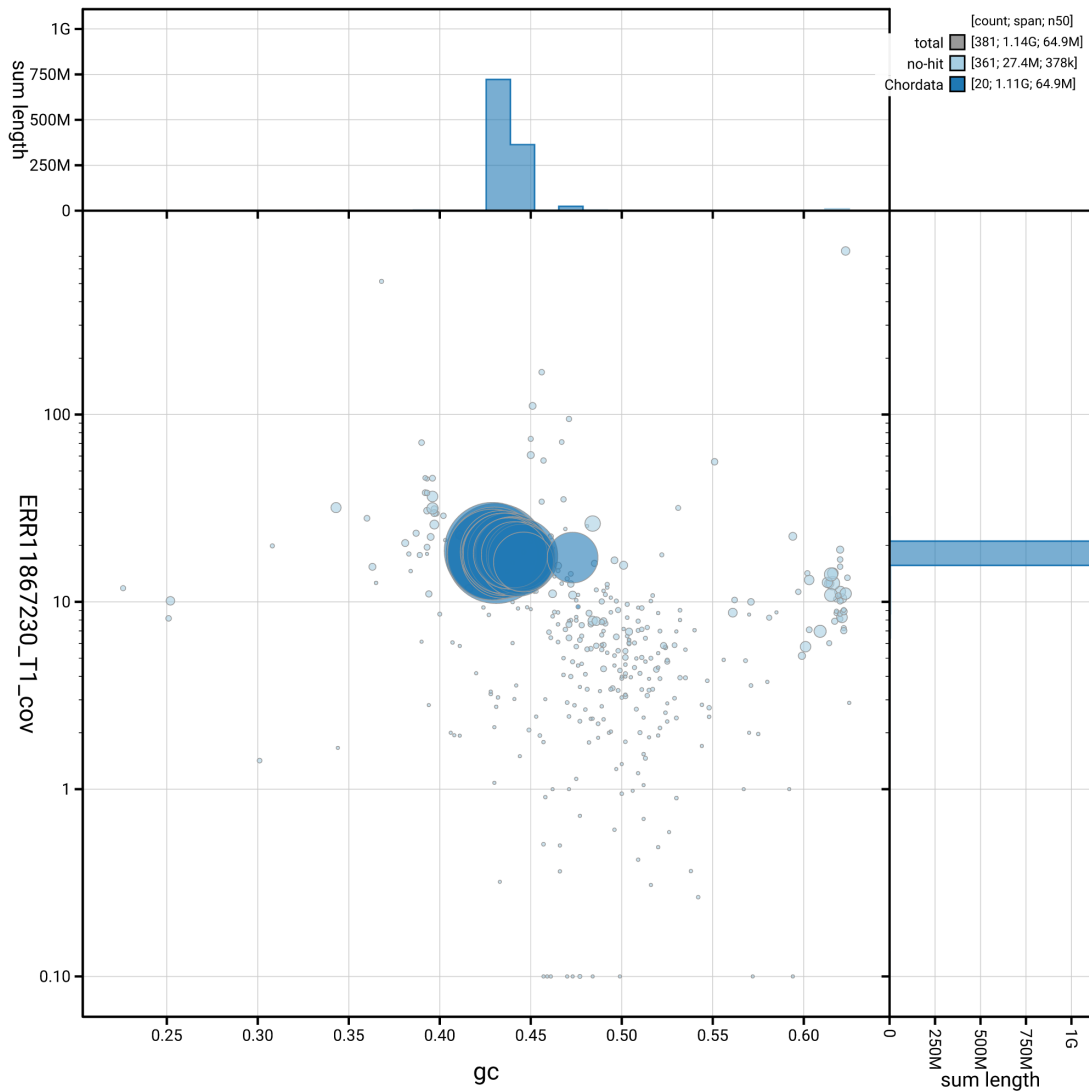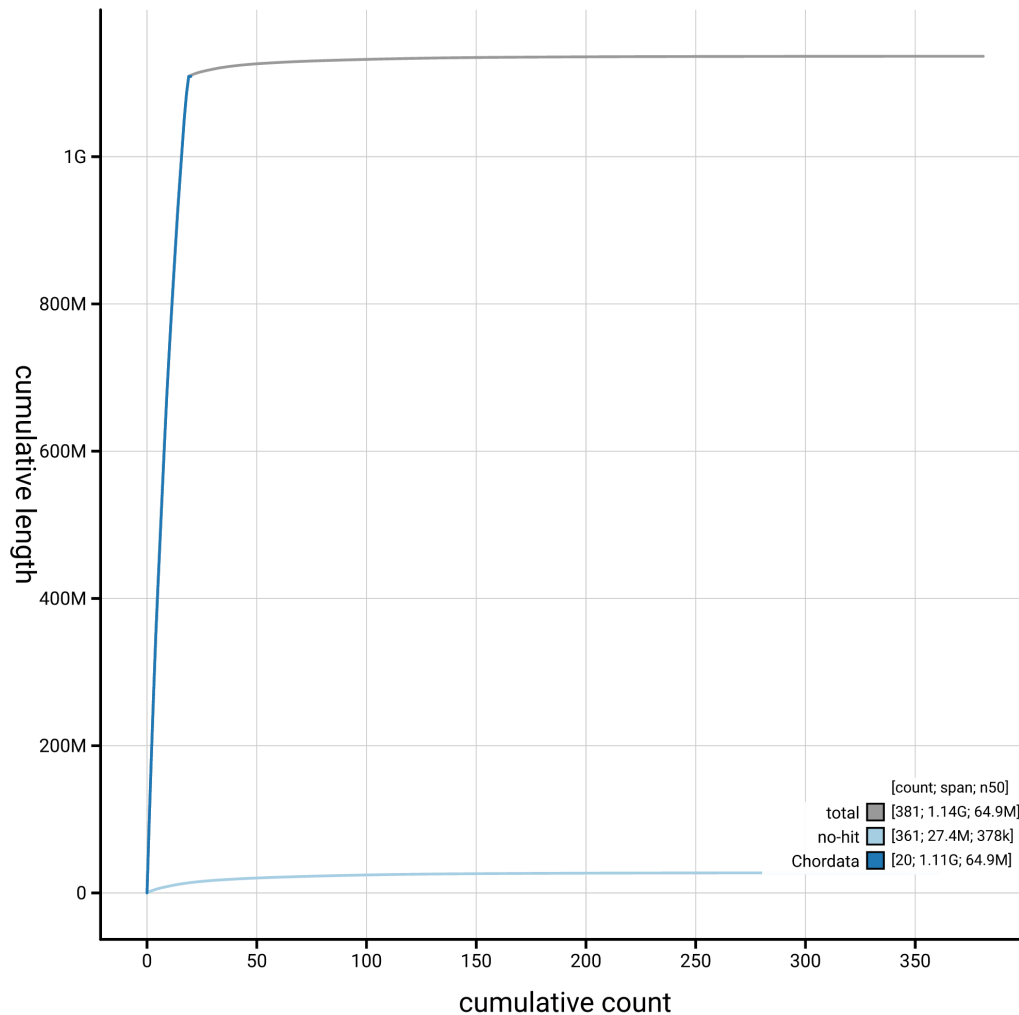
Figure 3. Genome assembly of *Conger conger*, fConCon1.1: BlobToolKit GC-coverage plot. Sequences are coloured by phylum. Circles are sized in proportion to sequence length. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963514075.1/dataset/GCA_963514075.1/blob.

a sequence of core procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b).

The fConCon1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). Tissue from the gill 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 (Todorovic *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using

AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Strickland *et al.*, 2023). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. The fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from muscle tissue of fConCon1 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

Figure 4. Genome assembly of *Conger conger* fConCon1.1: 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/GCA_963514075.1/dataset/GCA_963514075.1/cumulative.

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.

Library preparation and sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed 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 IIe (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments.

Hi-C data were generated from frozen gill tissue of the fConCon11 sample, using the Arima-HiC v2 kit. The tissue was fixed with a TC buffer containing formaldehyde, resulting

in crosslinked DNA. The crosslinked DNA was digested with a restriction enzyme master mix. The resulting 5'-overhangs were filled in and labelled with a biotinylated nucleotide. The biotinylated DNA was then fragmented, enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit. Hi-C sequencing was performed on an Illumina NovaSeq 6000 instrument, using paired-end sequencing with a read length of 150 bp.

Genome assembly, curation and evaluation
Assembly
The HiFi reads were first 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 were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were
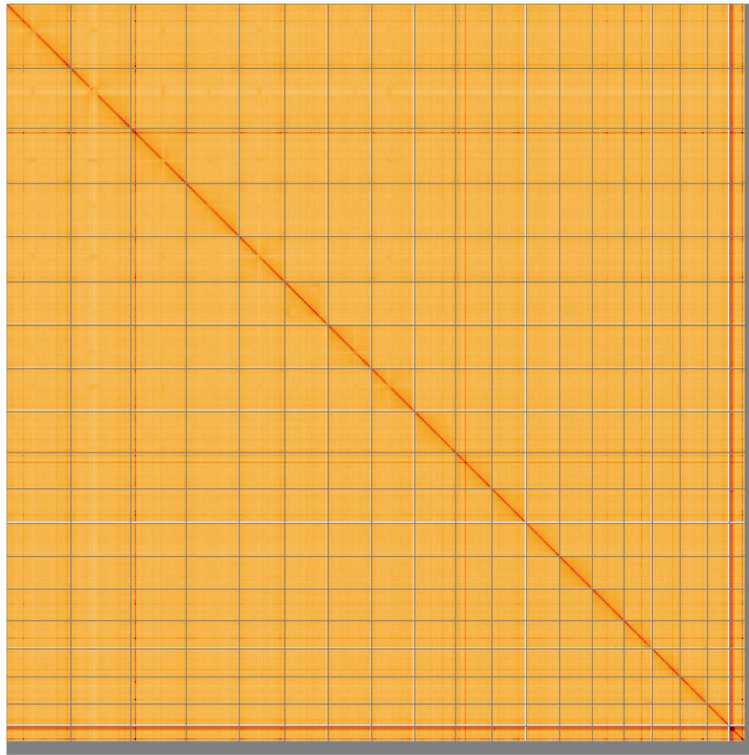
Figure 5. Genome assembly of *Conger conger* fConCon1.1: Hi-C contact map of the fConCon1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=Wd1d4hWJSj-yRmTB5eV_PA.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Conger conger*, fConCon1.

INSDC accession	Name	Length (Mb)	GC%
OY741314.1	1	96.78	43.0
OY741315.1	2	89.85	43.0
OY741316.1	3	83.29	43.0
OY741317.1	4	79.47	43.0
OY741318.1	5	68.37	43.5
OY741319.1	6	65.24	43.5
OY741320.1	7	65.18	43.5
OY741321.1	8	64.88	44.0
OY741322.1	9	61.47	43.5
OY741323.1	10	54.23	44.5
OY741324.1	11	51.85	44.0
OY741325.1	12	50.02	44.0
OY741326.1	13	48.9	44.0
OY741327.1	14	47.67	44.0
OY741328.1	15	42.41	44.5

INSDC accession	Name	Length (Mb)	GC%
OY741329.1	16	41.91	44.5
OY741330.1	17	40.83	44.0
OY741331.1	18	33.25	44.5
OY741332.1	19	23.41	47.5
OY741333.1	MT	0.02	37.0

further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option. 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 (article in preparation). Flat files and maps used in curation were

generated in TreeVal (Pointon *et al.*, 2023). 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 curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation (article in preparation).

Evaluation of the final assembly

The final assembly was post-processed and evaluated with the three Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines "sanger-tol/readmapping" (Surana *et al.*, 2023a), "sanger-tol/genomenote" (Surana *et al.*, 2023b), and "sanger-tol/blobtoolkit" (Muffato *et al.*, 2024). The pipeline sanger-tol/readmapping aligns the Hi-C reads with bwa-mem2 (Vasimuddin *et al.*, 2019) and combines the alignment files with SAMtools (Danecek *et al.*, 2021). The sanger-tol/genomenote pipeline transforms the Hi-C alignments into a contact map with BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020), which is then visualised with HiGlass (Kerpedjiev *et al.*, 2018). It also provides statistics about the assembly with the NCBI datasets (Sayers *et al.*, 2024) report, computes *k*-mer completeness and QV consensus quality values with FastK and MERQURY.FK, and a completeness assessment with BUSCO (Manni *et al.*, 2021).

The sanger-tol/blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads with SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoaT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineage, the pipeline aligns the BUSCO genes to the Uniprot Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND (Buchfink *et al.*, 2021) blastp. The genome is also split into chunks according to the density of the BUSCO genes from the closest taxonomically lineage, and each chunk is aligned to the Uniprot Reference Proteomes database with DIAMOND blastx. Genome sequences that have no hit are then chunked with seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). All those outputs are combined with the blobtools suite into a blobdir for visualisation.

The genome assembly and evaluation 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

Table 4. Software tools: versions and sources.

Software tool	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	4.3.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.4.3 and 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	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	https://github.com/thegenemyers/FASTK
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	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
Merqury.FK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK

Software tool	Version	Source
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets
Nextflow	23.04.0-5857	https://github.com/nextflow-io/nextflow
PretextView	0.2	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.16.1, 1.17, and 1.18	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/genomenote	1.1.1	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.2.1	https://github.com/sanger-tol/readmapping
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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)

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: Conger conger (European conger). Accession number PRJEB65259; https://identifiers.org/ena.embl/PRJEB65259 (Wellcome Sanger Institute, 2023).

The genome sequence is released openly for reuse. The *Conger conger* 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 Marine Biological Association Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.8382513.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.12158331

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.12162482.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.12165051.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.12160324.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.12205391.

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

References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* Oxford University Press, 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.* Blackwell Publishing Ltd, 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**

Bauchot ML, Blache J: **Presence d'Ariosoma balearicum (de la Roche, 1809) en mer Rouge (Pisces, Teleostei, Congridae). 1980.** *Bulletin Du Muséum National d'histoire Naturelle. Section A Zoologie, Biologie et Écologie Animales.* 1980; **4**(1): 1131–1137.

Beasley J, Uhl R, Forrest LL, *et al.*: **DNA barcoding SOPs for the Darwin Tree of Life project.** *protocols.io.* 2023.
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**

Cau A, Manconi P: **Sex-ratio and spatial displacement in *Conger conger*(L.).** *Rapp P-V Reun Ciesm.* 1983; **28**(5): 93–96.
Reference Source

Cau A, Manconi P: **Relationship of feeding, reproductive cycle and bathymetric distribution in *Conger conger*.** *Mar Biol.* 1984; **81**: 147–151.
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] Missing].** *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).* Genetics Society of America, 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.* Nature Research, 2021; **18**(2): 170–175.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Correia A, Isidro E, Antunes C, *et al.*: **Age growth, distribution and ecological aspects of *Conger conger* leptocephali collected in the Azores, based on otolith analysis of premetamorphic specimens.** *Mar Biol.* 2002; **141**: 1141–1151.
Publisher Full Text

Correia AT, Barros F, Sial AN: **Stock discrimination of European conger eel (*Conger conger* L.) using otolith stable isotope ratios.** *Fish Res.* 2011; **108**(1): 88–94.
Publisher Full Text

Correia AT, Manso S, Coimbra J: **Age growth and reproductive biology of the European conger eel (*Conger conger*) from the Atlantic Iberian waters.** *Fish Res.* 2009; **99**(3): 196–202.
Publisher Full Text

Correia AT, Ramos AA, Barros F, *et al.*: **Population structure and connectivity of the European conger eel (*Conger conger*) across the north-eastern Atlantic and western Mediterranean: integrating molecular and otolith elemental approaches.** *Mar Biol.* 2012; **159**(7): 1509–1525.
Publisher Full Text

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

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

da Veiga Leprevost F, Grüning BA, Aflitos SA, *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

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**

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**

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**

Fannon E, Fahy E, O'Reilly R: **Maturation in female congereel *Conger conger* (L.).** *J Fish Biol.* 1990; **36**(2): 275–276.
Publisher Full Text

Figueiredo M, Figueiredo I, Correia J: **Caracterizacao geral dos recursos de profundidade em estudo no IPIMAR [Instituto Portugues de Investigacao Maritima].** Relatorios Cientificos e Tecnicos (Portugal). No. 21. 1996.
Reference Source

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

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.* Oxford University Press, 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.
Reference Source

Hayward PJ, Ryland JS: **Handbook of the Marine Fauna of North-West Europe (2nd edn).** Oxford University Press, 2017.
Publisher Full Text

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**

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.* BioMed Central Ltd, 2018; **19**(1): 125.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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**

Levy A, Able KW, Grimes CB, *et al.*: **Biology of the conger eel *Conger oceanicus* in the Mid-Atlantic Bight.** *Mar Biol.* 1988; **98**: 597–600.
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**

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.* Oxford University Press, 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.
Reference Source

Muffato M, Butt Z, Challis R, *et al.*: **Sanger-tol/blobtoolkit: v0.3.0 – poliwag.** 2024.
Publisher Full Text

Pointon DL, Eagles W, Sims Y, *et al.*: **Sanger-tol/treeval v1.0.0 – Ancient Atlantis.** 2023.
Publisher Full Text

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
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.* Cell Press, 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.* Nature Research, 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.* BioMed Central Ltd, 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Saldanha L, Almeida AJ, Andrade F, *et al.*: **Observations on the diet of some slope dwelling fishes of Southern Portugal.** *Internationale Revue Der Gesamten Hydrobiologie Und Hydrographie.* 1995; **80**(2): 217–234.
Publisher Full Text

Sayers EW, Cavanaugh M, Clark K, *et al.*: **GenBank 2024 update.** *Nucleic Acids Res.* 2024; **52**(D1): D134–D137.
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

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

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

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: awaiting peer review].** *Wellcome Open Res.* 2024; **9**: 339.
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

Wellcome Sanger Institute: **The genome sequence of the European conger eel, *Conger conger* (Linnaeus, 1758).** European Nucleotide Archive. [dataset], accession number PRJEB65259, 2023.

Wheeler A: **World encyclopedia of fishes.** Macdonald, London, 1985.
Reference Source

Whitehead PJP: **Fishes of the North-Eastern Atlantic and the Mediterranean.** 1985.

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

Open Peer Review

Current Peer Review Status: ✓ ✓ ✓

Version 1

Reviewer Report 15 October 2024

https://doi.org/10.21956/wellcomeopenres.25388.r101469

✓ **Kazuharu Nomura**

Fisheries Technology Institute, Japan Fisheries Research and Education Agency, Mie, Japan

This manuscript is a Data Note type of submission that provides a chromosome-level genome assembly of the European conger eel (*Conger conger*). The paper is overall very clear, with detailed descriptions of the pipelines used to generate the dataset. The protocols are appropriate and technically sound, and the accuracy of the resulting genome assembly is reasonable according to current standards. All data and tools have been made publicly available, allowing other researchers easy access to this valuable resource. The overall quality of the research is high, and the provided data will be useful for future studies in the conservation and management of this species. Therefore, I would recommend that this paper be considered for indexing in Wellcome Open Research.

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: 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.

Reviewer Report 15 October 2024

https://doi.org/10.21956/wellcomeopenres.25388.r101468

✔ **Khairul Syahputra**
Research Center for Fishery, National Research and Innovation Agency (BRIN), Cibinong, Indonesia

Comments on manuscript of European conger genome:

The present study is an interesting work providing genome sequence of European conger. It can be useful for further studies related to genome references in Eel and other aquatic species. However, some minor issues exist in the text. They should be addressed before the manuscript is suitable for indexing.

Title:
"eel" is not necessary and can be omitted!

Abstract:
Common name of species should come first before its scientific name. I do not think that the classification of species deserves to be stated here! We can find it on the Species Taxonomy section!

Background:
As suggested previously, "eel" on common name of species is not necessarily included.
Line 5 – 500 m???
7 – "eel" or Eels"??
11 -2.7 m in length???
Second paragraph – Justification of using European eel as fish model in this study deserves more explanation here and why this species is focused on the present study!

Genome sequence report:
Line 1 – use a common name to mention the species. The scientific name is only included at the first appearance of species in the text. Apply it thorough manuscript!
Paragraph 2 line 3 – "1,009 gaps" is not found in Table2!

Methods:
Sample acquisition and barcoding
First paragraph, last sentence – a) What were tissue samples collected, and b) How long were the tissue sample kept on dry ice before nucleic acid extraction? State this!
2nd Paragraph line 3 – What tissue? Is there any specific size or weight?

Library preparation and Sequencing:
Why did the authors use DNA and RNA to build the genome sequence of European eel? Explain

this in the text!

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?
Partly

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Fish Genetic

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

Reviewer Report 23 September 2024

https://doi.org/10.21956/wellcomeopenres.25388.r100177

✅ **Chong Han**
Guangzhou University, Guangzhou, Guangdong, China

This study completed a chromosome-level assembly of European conger eel based on pacbio and Hi-C technology. It is a good work with detailed methods. All the protocols were appropriate and technically sound. Besides, we could find the related data in the NCBI database. In all, it is a valuable research, which is worthy to be Indexed in Wellcome Open Research.

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: endocrine disturbs, genome assembly and transcriptome

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.