



DATA NOTE

The genome sequence of the Hagfish, *Myxine glutinosa*

Linnaeus, 1758 (Myxiniformes: Myxinidae)

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

Anders Hay-Schmidt ¹,
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

¹Department of Odontology, University of Copenhagen, Copenhagen, Denmark

v1 First published: 03 Oct 2025, 10:537
<https://doi.org/10.12688/wellcomeopenres.24925.1>
Latest published: 03 Oct 2025, 10:537
<https://doi.org/10.12688/wellcomeopenres.24925.1>

Abstract

We present a genome assembly from an individual *Myxine glutinosa* (Hagfish; Chordata; Myxini; Myxiniformes; Myxinidae). The genome sequence has a total length of 3 057.50 megabases. Most of the assembly (87.13%) is scaffolded into 14 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 19.65 kilobases.

Keywords

Myxine glutinosa; Hagfish; genome sequence; chromosomal; Myxiniformes



This article is included in the [Tree of Life gateway](#).

Open Peer Review

Approval Status

? ✓ ✓

	1	2	3
version 1			
03 Oct 2025	view	view	view
1. Ferdinand Marlétaz  , University College London, London, UK			
2. Oliver Berry  , CSIRO, Crawley, Australia			
3. Ivan Rodrigo Wolf  , UNC Charlotte Department of Biological Sciences, Charlotte, 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: Hay-Schmidt A: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award (218328).

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award (218328).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Hay-Schmidt A *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Hay-Schmidt A, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations *et al.* **The genome sequence of the Hagfish, *Myxine glutinosa* Linnaeus, 1758 (Myxiniiformes: Myxinidae) [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2025, 10:537 <https://doi.org/10.12688/wellcomeopenres.24925.1>

First published: 03 Oct 2025, 10:537 <https://doi.org/10.12688/wellcomeopenres.24925.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Cyclostomata; Myxini; Myxiniformes; Myxinidae; Myxinae; *Myxine*; *Myxine glutinosa* Linnaeus, 1758 (NCBI:txid7769)

Background

Myxine glutinosa (Figure 1), commonly known as the Atlantic hagfish, was first described by Linnaeus in 1758 under *Vermes* – alongside earthworms – before its true affinity with jawless vertebrates (*Cyclostomi*) was recognised (Krøyer, 1852–1853; Linnaeus, 1758). The genus *Myxine* contains approximately 24 species, three of which (*M. glutinosa*, *M. ios* and *M. jespersenae*) occur in the North-East Atlantic. The exact relationships within *Myxine* remain unresolved: *M. glutinosa* appears to comprise an eastern clade (Greenland populations, *M. glutinosa glutinosa*) and a western clade (sometimes treated as *M. glutinosa limnosa*) (Fernholm *et al.*, 2013; Jensen, 1941). DNA analyses support recognising two species (Fernholm *et al.*, 2013), yet clear morphological characters are lacking (Møller *et al.*, 2005). The *Cyclostomi* represent the sister group to jawed vertebrates (*Gnathostomi*); obtaining chromosome-level hagfish genomes is therefore essential for reconstructing the early genomic events that separate cyclostome and gnathostome lineages (Marlétaz *et al.*, 2024; Yu *et al.*, 2024).

An adult Atlantic hagfish has an eel-shaped, cartilage-supported body, 15–45 cm long (occasionally up to 95 cm) and 1–1.5 cm in diameter, tapering towards the tail. Its smooth skin lacks scales and a lateral line but bears 88–111 mucus-gland openings per side, which produce copious slime for defence and burrow support (Fernholm & Vladykov, 1984; Møller *et al.*, 2005). The head carries a single nostril, four sensory barbels around a ventral mouth, and a rasp-like dental apparatus of two cartilages bearing some 30–38 teeth plus a median tooth. Beneath the skin lie rudimentary, light-sensitive eyes and scattered photoreceptors (Gustafson, 1935; Newth & Ross, 1955). Hagfish respire via six (sometimes seven) gill pouches that share one posterior opening per side; water is drawn through the nostril into the pharynx by a rapidly beating velum (Martini & Flescher, 2002; Otterstrøm, 1917; Pfaff, 1950).

Atlantic hagfish occupy soft-bottom habitats from about 27 to over 900 m depth, at temperatures of 1.5–10 °C and salinities above ca 31 ‰; they tolerate short periods at 29–31 ‰ but perish below 20–25 ‰ (Bigelow & Schroeder, 1948; Møller *et al.*, 2005). They live in slime-lined burrows, emerge to scavenge carrion or capture live prey, and may remain in low-salinity waters for weeks. Reproduction is dioecious and likely external; sexual maturity is reached at roughly 25 cm length and three years of age (Breder & Rosen, 1966;

van der Meer & Kooijman, 2014). Oocytes measure 20–25 mm, attach by anchor fibres, and develop slowly; larval stages remain undocumented (Holmgren, 1946; Martini, 2000).

As scavengers and bioturbators, *M. glutinosa* contributes to nutrient cycling and seabed oxygenation (Martini *et al.*, 1997). Predators include cod, dogfish, Greenland shark and harbour porpoise (Martini & Flescher, 2002; Nielsen *et al.*, 2014). The species is widespread and abundant – listed as Least Concern by the IUCN – though small-scale fisheries in the western Atlantic warrant monitoring (DFO, 2023; Keith, 2006).

Methods

Sample acquisition

The specimen used for genome sequencing was an adult *Myxine glutinosa* (specimen ID SAN00002398, ToLID kmMyxGlut1), collected from Kristineberg Marine Biological Station, Sweden (latitude 58.2508, longitude 11.4602) on 2021-06-14. The specimen was collected and identified by Anders Hay-Schmidt (University of Copenhagen). Sample metadata were collected in line with the Darwin Tree of Life project standards described by Lawniczak *et al.* (2022).

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The kmMyxGlut1 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the liver was homogenised by cryogenic disruption using the Covaris cryoPREP® Automated Dry Pulverizer. HMW DNA was extracted using the Automated MagAttract v2 protocol. DNA was sheared into an average fragment size of 12–20 kb following the Megaruptor®3 for LI PacBio protocol. Sheared DNA was purified by automated SPRI (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 6.96 ng/μL and a yield of 2 714.40 ng. The 260/280 spectrophotometric ratio was 1.79, and the 260/230 ratio was 1.11.

RNA was extracted from liver tissue of kmMyxGlut1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana protocol. The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes reagents for end repair/A-tailing, adapter ligation, post-ligation



Figure 1. *Myxine glutinosa* (hagfish). Plate 120 from Georges Cuvier, *Le règne animal distribué d'après son organisation, seconde édition, Tome VIII* (1828).

SMRTbell bead clean-up, and nuclease treatment. Size selection and clean-up were performed using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using a Qubit Fluorometer v4.0 (ThermoFisher Scientific) and the Qubit 1X dsDNA HS assay kit. Final library fragment size was assessed with the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit.

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

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen whole organism tissue of the kmMyxGlut1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again and equimolar

and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq 6000.

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo(dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq 6000 to generate 150-bp paired-end reads.

Genome assembly

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

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with the --break flag for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

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

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 168 breaks, 407 joins, and removal of 2 haplotypic duplications. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextView-Snapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

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

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

management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Myxine glutinosa* specimen generated 88.17 Gb (gigabases) from 9.99 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 2 719.74 Mb, with a heterozygosity of 1.19% and repeat content of 72.34% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 30× coverage. Hi-C sequencing produced 896.44 Gb from 5 936.70 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The final assembly has a total length of 3 057.50 Mb in 4 003 scaffolds, with 4 412 gaps, and a scaffold N50 of 192.17 Mb (Table 2).

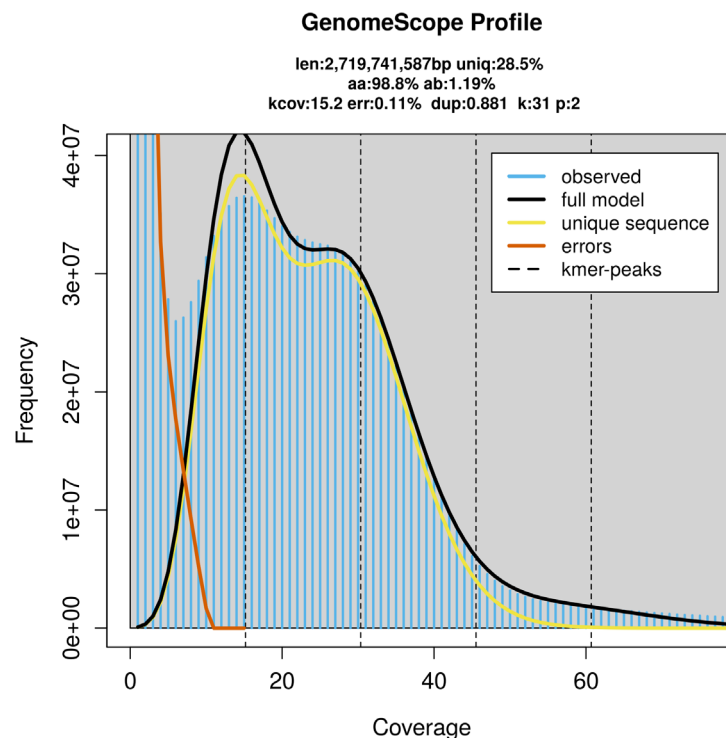


Figure 2. Frequency distribution of k -mers generated using GenomeScope2. The plot shows observed and modelled k -mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB69502.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	kmMyxGlut1	kmMyxGlut1	kmMyxGlut1
Specimen ID	SAN00002398	SAN00002398	SAN00002398
BioSample (source individual)	SAMEA12790223	SAMEA12790223	SAMEA12790223
BioSample (tissue)	SAMEA12790224	SAMEA12790223	SAMEA12790224
Tissue	liver	whole organism	liver
Sequencing platform and model	Sequel IIE	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR12303934; ERR12303935; ERR12303937; ERR12303933; ERR12303936	ERR12318581; ERR12318583	ERR12318582
Read count total	9.99 million	5 936.70 million	65.53 million
Base count total	88.17 Gb	896.44 Gb	9.89 Gb

Table 2. Genome assembly statistics.

Assembly name	kmMyxGlut1.1
Assembly accession	GCA_964187855.1
Alternate haplotype accession	GCA_964187825.1
Assembly level	chromosome
Span (Mb)	3 057.50
Number of chromosomes	14
Number of contigs	8 415
Contig N50	0.74 Mb
Number of scaffolds	4 003
Scaffold N50	192.17 Mb
Sex chromosomes	N/A
Organelles	Mitochondrial genome: 19.65 kb

Most of the assembly sequence (87.13%) was assigned to 14 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

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

The combined primary and alternate assemblies achieve an estimated QV of 54.5. The *k*-mer completeness is 84.20% for

the primary assembly, 81.95% for the alternate haplotype, and 95.75% for the combined assemblies (Figure 4).

BUSCO v.5.5.0 analysis using the vertebrata_odb10 reference set (*n* = 3 354) identified 57.0% of the expected gene set (single = 54.9%, duplicated = 2.1%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage.

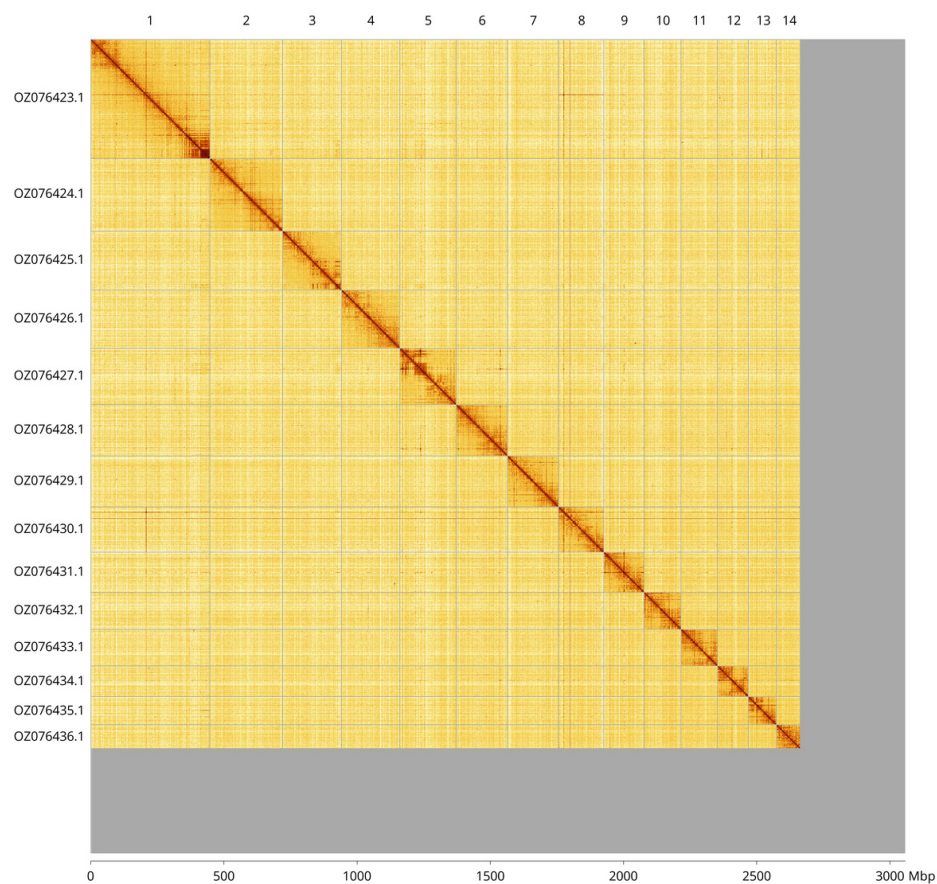


Figure 3. Hi-C contact map of the *Myxine glutinosa* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Myxine glutinosa* kmMyxGlut1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ076423.1	1	448.54	45
OZ076424.1	2	272.60	45
OZ076425.1	3	220.90	44.50
OZ076426.1	4	218.61	45
OZ076427.1	5	212.07	45
OZ076428.1	6	192.17	45
OZ076429.1	7	191.53	45
OZ076430.1	8	169.62	45
OZ076431.1	9	151.20	45
OZ076432.1	10	138.58	45
OZ076433.1	11	138.12	45
OZ076434.1	12	114.45	45
OZ076435.1	13	106.09	45
OZ076436.1	14	89.40	45

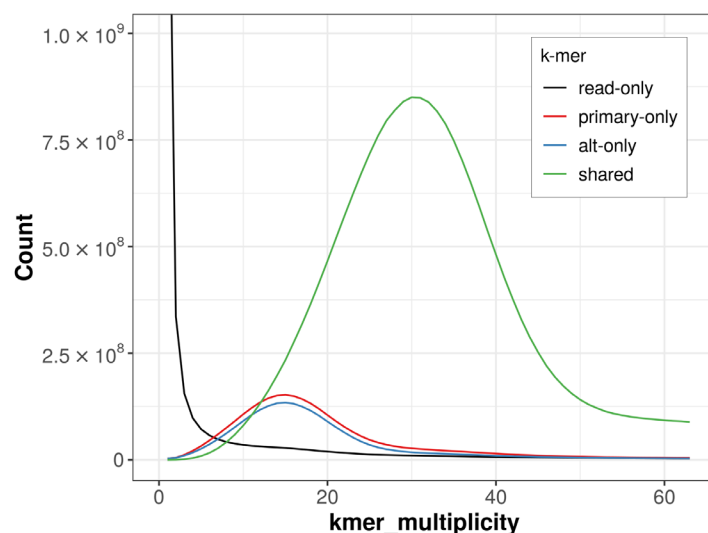


Figure 4. Evaluation of k -mer completeness using MerquyFK. This plot illustrates the recovery of k -mers from the original read data in the final assemblies. The horizontal axis represents k -mer multiplicity, and the vertical axis shows the number of k -mers. The black curve represents k -mers that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

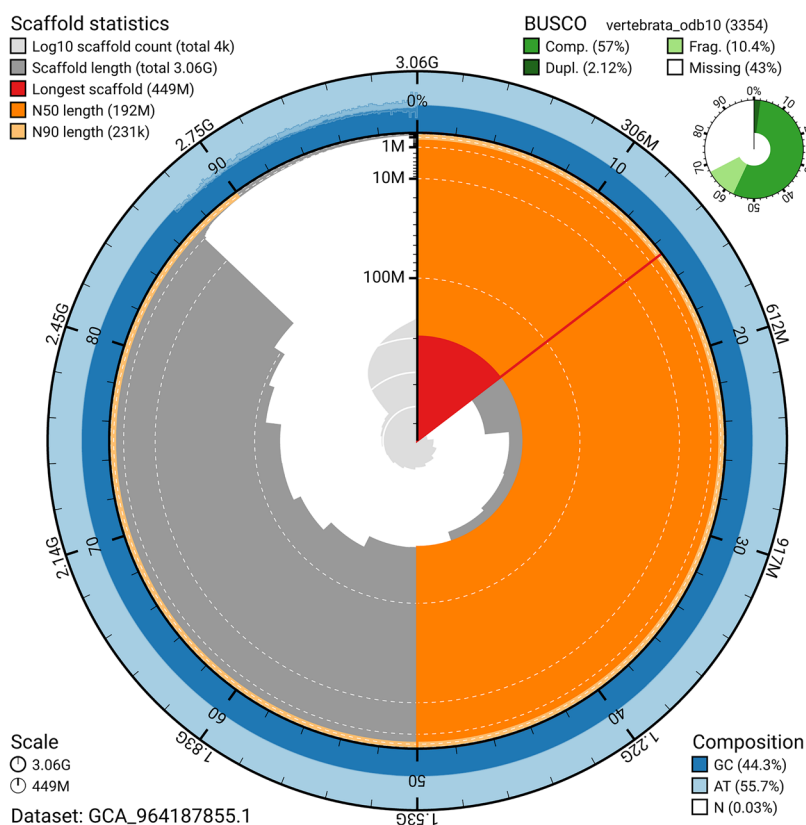


Figure 5. Assembly metrics for kmMyxGlut1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the vertebrata_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the [BlobToolKit viewer](#).

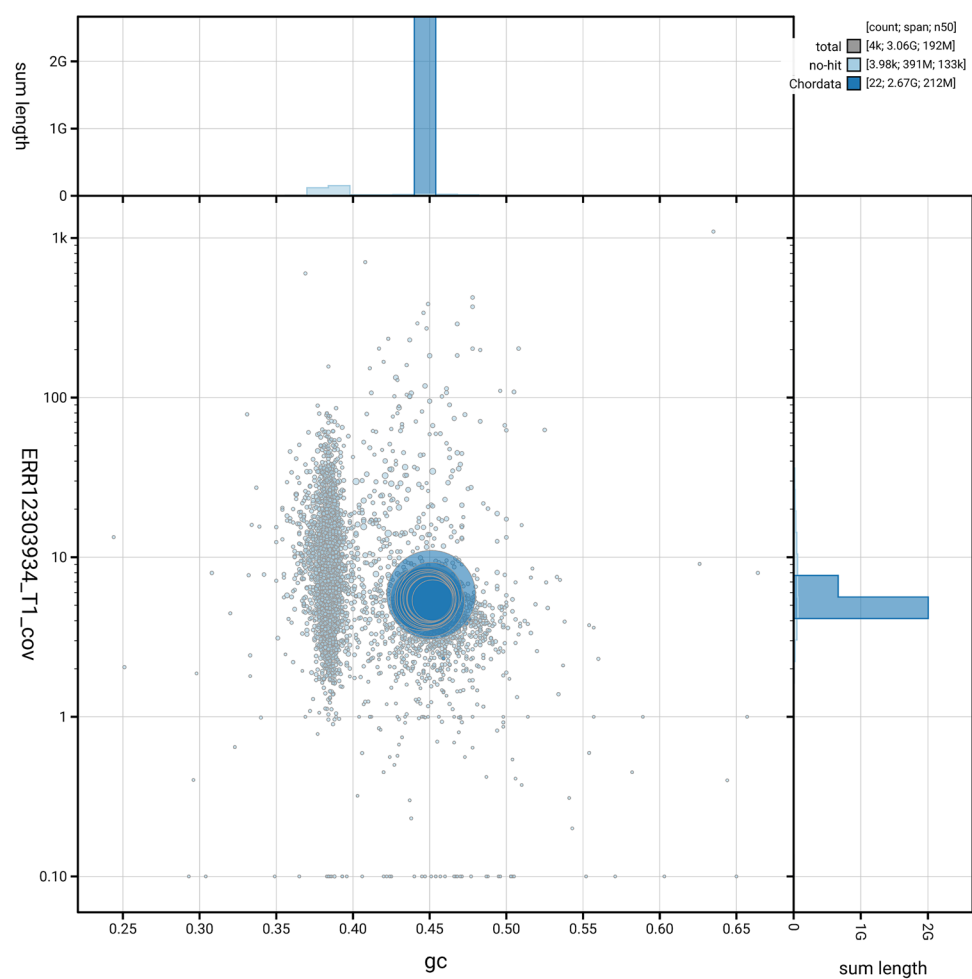


Figure 6. BlobToolKit GC-coverage plot for kmMyxGlut1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the [BlobToolKit viewer](#).

Table 4. Earth Biogenome Project summary metrics for the *Myxine glutinosa* assembly.

Measure	Value	Benchmark
EBP summary (primary)	5.8.Q53	6.C.Q40
Contig N50 length	0.74 Mb	≥ 1 Mb
Scaffold N50 length	192.17 Mb	= chromosome N50
Consensus quality (QV)	Primary: 53.8; alternate: 55.0; combined: 54.5	≥ 40
<i>k</i> -mer completeness	Primary: 84.20%; alternate: 81.95%; combined: 95.75%	≥ 95%
BUSCO	C:57.0% [S:54.9%; D:2.1%]; F:10.4%; M:32.6%; n:3 354	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	87.13%	≥ 90%

Table 4 lists the assembly metric benchmarks adapted from Rhie *et al.* (2021) the Earth BioGenome Project Report on Assembly Standards September 2024. The EBP metric, calculated for the primary assembly, is 5.8.Q53.

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 web-site. 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: *Myxine glutinosa* (Atlantic hagfish). Accession number PRJEB69502. The genome sequence is released openly for reuse. The *Myxine glutinosa* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665), the Sanger Institute Tree of Life Programme (PRJEB43745) and the Vertebrate Genomes Project (PRJNA489243). All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

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

Author information

- Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team
- Members of Wellcome Sanger Institute Scientific Operations – Sequencing Operations
- Members of the Wellcome Sanger Institute Tree of Life Core Informatics team
- Members of the Tree of Life Core Informatics collective
- Members of the Darwin Tree of Life Consortium

Table 5. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
fasta_windows	0.2.4	https://github.com/tolkkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats

Software	Version	Source
Goat CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MercuryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0	https://github.com/nextflow-io/nextflow
PretextView	N/A	https://github.com/sanger-tol/PretextView
PurgeDups	1.2.5	https://github.com/dfguan/purge_dups
Samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

- Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bigelow HB, Schroeder WC: **Cyclostomes.** In: *Fishes of the Western North Atlantic. Part I.* New Haven, USA: Sears Foundation for Marine Research, 1948; 29–58.
- Breder CM, Rosen DE: **Modes of reproduction in fishes.** Garden City, New York, USA: Natural History Press, 1966.
- 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](#)
- 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](#)
- 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](#)
- DFO: **Decision on Atlantic hagfish fishery management.** 2023.
- 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](#)
- Fernholm B, Norén M, Kullander SO, *et al.*: **Hagfish phylogeny and taxonomy, with description of the new genus *Rubicundus* (craniata, myxinidae).** *J Zool Syst Evol Res.* 2013; **51**(4): 296–307.
[Publisher Full Text](#)
- Fernholm B, Vladikov VD: **Myxinidae.** In: P. J. P. Whitehead, M.-L. Bauchot, J.-C. Hureau, *et al.* (eds), *Fishes of the North-eastern Atlantic and the Mediterranean, Volume I.* UNESCO, 1984; 68–69.
- 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.* 2020; **36**(9): 2896–2898.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gustafson G: **On the biology of *Myxine glutinosa* L.** *Arkiv för Zoologi.* 1935;

28A(2): 1–8.

Holmgren N: **On two embryos of *Myxine glutinosa*.** *Acta Zool.* 1946; **27**(1): 1–90.

[Publisher Full Text](#)

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

[Publisher Full Text](#)

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): g1aa153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Jensen AS: **The marsipobranchs of Greenland.** *Videnskabelige Meddelelser fra Dansk Naturhistorisk Forening, København.* 1941; **105**: 55–57.

Keith C: **Status of fishery resources off the northeastern US (Atlantic hagfish, *Myxine glutinosa*).** 2006.

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](#)

Krøyer HN: **Danske Fiske Vol. 3 Part 2.** København: S. Triers Officin, 1852–1853.

[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](#)

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

[Publisher Full Text](#)

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Linnaeus C: **Systema Naturae, 10, Tomus 1. Pp. 641, 644, and 650.** 1758.

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](#)

Marlétaz F, Timoshevskaya N, Timoshevskiy VA, *et al.*: **The hagfish genome and the evolution of vertebrates.** *Nature.* 2024; **627**(8005): 811–820.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Martini FH: **The evidence for and potential significance of cutaneous respiration in hagfishes.** *FASEB J.* 2000; **14**(4): A436.

Martini FH, Flescher D: **Hagfishes. Order myxiniiformes.** In: B. B. Collette and G. Klein-MacPhee (eds), *Bigelow & Schroeder's Fishes of the Gulf of Maine.* Smithsonian Institution Press, 2002; 9–16.

[Reference Source](#)

Martini FH, Lesser MP, Heiser JB: **Ecology of the hagfish, *Myxine glutinosa* L, in the gulf of Maine: II. Potential impact on benthic communities and commercial fisheries.** *J Exp Mar Biol Ecol.* 1997; **214**(1–2): 97–106.

[Publisher Full Text](#)

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

[Reference Source](#)

Møller PR, Feld TK, Poulsen IH, *et al.*: ***Myxine jespersenae*, a new species of hagfish (myxiniiformes: Myxiniidae) from the North Atlantic Ocean.** *Copeia.* 2005; **2005**(2): 374–85.

[Publisher Full Text](#)

Newth DR, Ross DM: **On the reaction to light of *Myxine glutinosa* L.** *J Exp Biol.* 1955; **32**(1): 4–21.

[Publisher Full Text](#)

Nielsen J, Hedeholm RB, Simon M, *et al.*: **Distribution and feeding ecology of the Greenland shark (*Somniosus microcephalus*) in Greenland waters.** *Polar Biol.* 2014; **37**: 37–46.

[Publisher Full Text](#)

Otterstrøm CV: **Danmarks Fauna Bd. 20. Fisk III, Fastkæbede, Buskgællede, Ganoider, Tværmunde Og Rundmunde.** København: G. E. C. Gads Forlag, 1917.

Pfaff JR: **Slimålernes orden (hyperotreti).** In: F. W. Brædstrup, G. Thorson, and E. Wesenberg-Lund (eds), *Vort Lands Dyreliv. Andet Bind. Fisk, Hvirvelløse Dyr, Urdyr.* Gyldendalske Boghandel – Nordisk Forlag, 1950; 15–17.

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: a comprehensive update on curation, resources and tools.** *Database (Oxford).* 2020; **2020**: baaa062.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Uliano-Silva M, Ferreira JGRN, Krashenninnikova 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](#)

van der Meer J, Kooijman SALM: **Inference on energetics of deep-sea fish that cannot be aged: the case of the hagfish.** *J Sea Res.* 2014; **94**: 138–143.

[Publisher 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](#)

Yu D, Ren Y, Uesaka M, *et al.*: **Hagfish genome elucidates vertebrate whole-genome duplication events and their evolutionary consequences.** *Nat Ecol Evol.* 2024; **8**(3): 519–535.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:



Version 1

Reviewer Report 11 December 2025

<https://doi.org/10.21956/wellcomeopenres.27455.r139923>

© 2025 Wolf I. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ivan Rodrigo Wolf 

UNC Charlotte Department of Biological Sciences, Charlotte, North Carolina, USA

The manuscript describes the genome assembly of *Myxine glutinosa*, the Atlantic hagfish. *M. glutinosa* belongs to the order Cyclostomini, the sister group to *Gnathostomi* (the jawed vertebrates), and plays important ecological roles, being a scavenger that contributes to nutrient cycling and oxygenation of the seabed.

The biomolecule extraction and preparation protocols follow the high standards of the Wellcome Sanger Institute's (WSI) Tree of Life program, a program for investigating life using DNA sequencing, and appear suitable for generating a high-quality reference genome for the species.

I would just like to suggest that the authors clarify one point further: in the "Methods" section, under the "Genome assembly" first paragraph the authors state "(...) *was generated from the filtered reads using FastK*". However, no filtering methodology is mentioned at this point in the manuscript, a minor inconsistency that is easy to correct.

Overall, the genome assembly and quality metrics are excellent, given that differences in "k-mer completeness" and "Percentage of assembly assigned to chromosomes" values compared to the benchmark in Table 4 are expected.

The manuscript is suitable for indexed given its proposal to generate a reference genome for the species.

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: My specialization is in bioinformatics, with an emphasis on systems biology, genomics, transcriptomics, molecular phylogenetics, immunogenomics, pattern recognition, and the search and characterization of genes of biotechnological interest.

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

Reviewer Report 25 November 2025

<https://doi.org/10.21956/wellcomeopenres.27455.r139931>

© 2025 Berry O. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Oliver Berry 

CSIRO, Crawley, Australia

Overarching comments.

Thank you – a useful piece of work on a fascinating creature.

1. Although this is a data note, and thus largely about making a resource available, I felt that the authors could have provided a broader context for the value this particular resource represents. For example, indicating what other basal fish groups have genomes available, whether there are genomic features unique to these groups, whether provision of this genome is proposed to address specific knowledge gaps. This need not be detailed, but I believe would: a) improve readability; and b) strengthen the justification for the work.
2. Again, although it's a data note, some high-level reflection on the nature of the genome in relation to biological and evolutionary context and expectations would be welcomed by this reviewer. Similarly, I believe that some reflection on the assembly metrics versus benchmarks (Table 4) and how to interpret the generally lower values in relation to benchmarks would be valuable. Perhaps the differences are trivial. Perhaps it's clear to the authors how genome assembly metrics could be improved in the future. Again, these could be high level reflections, but I believe could help others in the field understand how to approach similar problems.

Key questions for a Data Note:

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

The rationale is briefly explained in terms of understanding early evolution of jawed fishes. I would have liked to have heard about existing genomic data relevant to this question, and where

provision of this genome would contribute.

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

I am not an expert on genome sequencing and assembly, but the analysis was thorough and provision of results adjacent to benchmark values was helpful.

- Are sufficient details of methods and materials provided to allow replication by others?
- Yes, they are thorough and links to additional information and resources are indicated.
- Are the datasets clearly presented in a useable and accessible format?

Yes, while presenting genome data can be challenging, given the scale, and richness, I appreciated the detailed figure captions supplied with figures.

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

Partly

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: Environmental genomics.

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 27 October 2025

<https://doi.org/10.21956/wellcomeopenres.27455.r135229>

© 2025 Marlétaz F. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ferdinand Marlétaz

Centre for Life's Origin & Evolution, Department of Ecology, Evolution & Environment, University College London, London, England, UK

This report describes the sequencing of the hagfish *Myxine glutinosa* genome using Pacbio Hifi technology and proximity ligation for scaffolding. This species is one of the available hagfish in Europe and its genome would constitute a valuable resource for research community. The genome assembly procedure is appropriately described and the datasets are accessible (though no

annotation was performed).

I however, have some reservations regarding this specific genome.

- It is the somatic and not the germline genome, while another genome for the same species with the germline has been reported (GCF_040869285.1)

- the coverage used (30x) appears relatively low to assemble with high confidence an heterozygous (1%) vertebrate genome. This reflects in the k-mer analysis, where the coverage seems insufficient to discriminate between erroneous and haplotype-related lower-frequency k-mers

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

Yes

Are the protocols appropriate and is the work technically sound?

Partly

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: evolution of animal genomes, evodevo, single cell

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
