DATA NOTE

The chromosomal genome sequence of the bigfin reef squid, *Sepioteuthis lessoniana* d'Orbigny, 1826 and its associated microbial metagenome sequences

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

Gustavo Sanchez[1,2], Oleg Simakov[3], Spencer Nyholm[4], Michele Nishiguchi[5], Margaret McFall-Ngai[6], Raphael Lami[7], Elizabeth Heath-Heckman[8], Graeme Oatley[9], Elizabeth Sinclair[9], Eerik Aunin[9], Noah Gettle[9], Camilla Santos[9], Michael Paulini[9], Haoyu Niu[9], Victoria McKenna[9], Rebecca O'Brien[9],
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,
EBI Aquatic Symbiosis Genomics Data Portal Team,
Aquatic Symbiosis Genomics Project Leadership

[1]Molecular Genetics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan
[2]Integrated Science for Life, Hiroshima University, Higashi Hiroshima, Hiroshima, Japan
[3]University of Vienna, Vienna, Vienna, Austria
[4]University of Connecticut, Storrs, Connecticut, USA
[5]University of California Merced, Merced, California, USA
[6]California Institute of Technology, Pasadena, California, USA
[7]Microbial Biodiversity et Biotechnology Lab, Sorbonne University, Banyuls-sur-mer, France
[8]Michigan State University, East Lansing, Michigan, USA
[9]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v1 **First published:** 14 Jul 2025, **10**:351
https://doi.org/10.12688/wellcomeopenres.24553.1
Latest published: 14 Jul 2025, **10**:351
https://doi.org/10.12688/wellcomeopenres.24553.1

Abstract

We present a genome assembly from a specimen of *Sepioteuthis lessoniana* (bigfin reef squid; Mollusca; Cephalopoda; Myopsida; Loliginidae). The genome sequence has a total length of 5,056.23 megabases. Most of the assembly (86.4%) is scaffolded into 44 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 16.64 kilobases in length. Gene annotation of this assembly on Ensembl identified 28,970 protein-coding genes.

Open Peer Review

Approval Status ✓ ✓ ?

	1	2	3
version 1 14 Jul 2025	✓ view	✓ view	? view

1. **Erich M. Schwarz**, Cornell University, Ithaca, USA

2. **David EK Ferrier**, University of St

Keywords
Sepioteuthis lessoniana, bigfin reef squid, genome sequence, chromosomal, Myopsida

This article is included in the Tree of Life gateway.

Andrews, Scotland, UK

3. **Ole Kristian Tørresen**, University of Oslo CEES (Ringgold ID: 206081), Oslo, Norway

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Aquatic Symbiosis Genomics Project Leadership (Mark.Blaxter@sanger.ac.uk)

Author roles: Sanchez G: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Simakov O**: Funding Acquisition, Resources, Supervision, Writing – Review & Editing; **Nyholm S**: Conceptualization, Funding Acquisition, Supervision; **Nishiguchi M**: Conceptualization, Funding Acquisition, Supervision; **McFall-Ngai M**: Conceptualization, Funding Acquisition, Supervision; **Lami R**: Conceptualization, Funding Acquisition, Supervision; **Heath-Heckman E**: Conceptualization, Funding Acquisition, Supervision; **Oatley G**: Investigation; **Sinclair E**: Investigation; **Aunin E**: Formal Analysis; **Gettle N**: Data Curation, Formal Analysis; **Santos C**: Formal Analysis; **Paulini M**: Formal Analysis; **Niu H**: Project Administration; **McKenna V**: Project Administration; **O'Brien R**: Project Administration;

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by the Gordon and Betty Moore Foundation through a grant (GBMF8897) to the Wellcome Sanger Institute to support the Aquatic Symbiosis Genomics Project, and by Wellcome through core funding to the Wellcome Sanger Institute (220540). GS received a 22K15085 Grant-in-Aid for Early-Career Scientists (KAKENHI).
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: Sanchez G, Simakov O, Nyholm S *et al.* **The chromosomal genome sequence of the bigfin reef squid, *Sepioteuthis lessoniana* d'Orbigny, 1826 and its associated microbial metagenome sequences [version 1; peer review: 2 approved, 1 approved with reservations]** Wellcome Open Research 2025, **10**:351 https://doi.org/10.12688/wellcomeopenres.24553.1

First published: 14 Jul 2025, **10**:351 https://doi.org/10.12688/wellcomeopenres.24553.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Spiralia; Lophotrochozoa; Mollusca; Cephalopoda; Coleoidea; Decapodiformes; Myopsida; Loliginidae; *Sepioteuthis*; *Sepioteuthis lessoniana* d'Orbigny, 1826 (NCBI:txid34570)

Background

The bigfin reef squid *Sepioteuthis lessoniana* Férussac, 1831 in Lesson (1830–1831) is a demersal neritic species that belongs the Order Myopsida and the Family Loliginidae. *S. lessoniana* is a species complex with three lineages distributed across the coast of the Indo-Pacific Ocean, from the western Indian Ocean in the Red Sea to the Central Pacific Ocean (Hawaii) (Cheng *et al.*, 2014). In Japanese waters, three cryptic species have been identified based on COI sequences, allozyme electrophoresis patterns and morphology, referred to as the "aka," "shiro," and "kua" types (Imai & Aoki, 2012; Tomano *et al.*, 2016). These types correspond to *Sepioteuthis* sp. 1, sp. 2, and sp. 3, respectively, which represent the three distinct lineages found throughout the entire distribution of the *S. lessoniana* sensu stricto. The "aka" and "shiro" types are abundant off mainland Japanese waters, while the "kua," "shiro," and "aka" types are all found in the waters of the Ryukyu Archipelago. In addition, the "kua" type is absent from mainland Japanese waters, while the "shiro" type is the most dominant species in mainland Japan.

Hatchlings of *Sepioteuthis* sp. 2 are larger and more developed than any other loliginids. This advanced development has enabled successful culturing of this species over several generations (Arnold, 1990; Shigeno *et al.*, 2001), the study of social behaviour such as schooling (Sakurai & Ikeda, 2024; Sugimoto *et al.*, 2013; Sugimoto & Ikeda, 2012) and camouflage (Nakajima *et al.*, 2022). A reference genome for this species will facilitate comparative genomics in the Myopsida Order, enable research into the neural mechanisms underlying social behaviour, and support selective breeding through genotyping, ultimately advancing aquaculture practices.

Genome sequence report

Sequencing data

The genome of a specimen of *Sepioteuthis lessoniana* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 249.90 Gb from 30.19 million reads. Based on the estimated genome size, the sequencing data provided approximately 34× coverage of the genome. Chromosome conformation Hi-C data produced 1,203.59 Gb from 7,970.82 million reads. RNA sequencing data were also gener-ated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing information.

Figure 1. Photograph of the *Sepioteuthis lessoniana* (xcSepLess1) specimen used for genome sequencing. (Photo Credits to Dr Satoshi Tomano.)

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The assembly was improved by manual curation, which corrected 577 misjoins or missing joins and removed 3 haplotypic duplications. These interventions reduced the total assembly length by 0.35%, decreased the scaffold count by 4.09%, and increased the scaffold N50 by 14.75%. The final assembly has a total length of 5,056.23 Mb in 6,819 scaffolds, with 7,097 gaps, and a scaffold N50 of 96.87 Mb (Table 2).

The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (86.38%) was assigned to 44 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3). During curation, it was noted that a large amount of repeat sequences could not be placed on chromosomes.

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 in GenBank.

Assembly quality metrics

The estimated Quality Value (QV) and *k*-mer completeness metrics, along with BUSCO completeness scores, were calculated for each haplotype and the combined assembly. The QV reflects the base-level accuracy of the assembly, while *k*-mer completeness indicates the proportion of expected *k*-mers identified in the assembly. BUSCO scores provide a measure of completeness based on benchmarking universal single-copy orthologues.

The combined primary and alternate assemblies achieve an estimated QV of 52.0. The *k*-mer completeness is 84.36% for the primary haplotype and 81.77% for the alternate haplotype; and 94.54% for the combined primary and alternate assemblies. BUSCO v.5.5.0 analysis using the mollusca_odb10 reference set (n = 5,295) identified 73.7% of the expected gene set (single = 72.7%, duplicated = 1.0%).

Metagenome report

Twenty-nine binned genomes were generated from the metagenome assembly, of which 17 were classified as high-quality

Table 1. **Specimen and sequencing data for *Sepioteuthis lessoniana*.**

Project information			
Study title	Sepioteuthis lessoniana (bigfin reef squid)		
Umbrella BioProject	PRJEB64979		
Species	*Sepioteuthis lessoniana*		
BioSpecimen	SAMEA11646235		
NCBI taxonomy ID	34570		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	xcSepLess1	SAMEA11646288	Somatic tissue
Hi-C sequencing	xcSepLess2	SAMEA11646318	Somatic tissue
RNA sequencing	xcSepLess2	SAMEA11646297	Gill
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR11837536	3.61e+09	545.23
Hi-C Illumina NovaSeq 6000	ERR11837535	4.36e+09	658.36
PacBio Sequel IIe	ERR11843445	2.75e+06	19.73
PacBio Sequel IIe	ERR11843447	2.69e+06	22.08
PacBio Sequel IIe	ERR11843448	2.63e+06	25.16
PacBio Sequel IIe	ERR11843449	2.64e+06	23.25
PacBio Sequel IIe	ERR11843451	2.88e+06	19.13
PacBio Revio	ERR12055552	8.23e+06	69.17
PacBio Sequel IIe	ERR11843446	2.94e+06	20.88
PacBio Sequel IIe	ERR11843450	2.93e+06	21.37
PacBio Sequel IIe	ERR11843452	2.49e+06	29.14
RNA Illumina NovaSeq X	ERR13093641	1.36e+08	20.58
RNA Illumina NovaSeq 6000	ERR11837537	4.42e+07	6.68

metagenome assembled genomes (MAGs) (see methods). The completeness values for these assemblies range from approximately 63% to 100% with contamination below 9%. A cladogram of the binned metagenomes is shown in Figure 6. For details on binned genomes see Table 4.

Genome annotation report

The *Sepioteuthis lessoniana* genome assembly (GCA_963585895.1) was annotated at the European Bioinformatics Institute (EBI) on Ensembl Rapid Release. The resulting annotation includes 68,971 transcribed mRNAs from 28,970 protein-coding and 21,583 non-coding genes (Table 2; https://beta.ensembl.org/species/f1e0d42d-d087-4e33-b0af-a3606c35ba3b). The average transcript length is 27,933.39, with 1.36 coding transcripts per gene and 4.95 exons per transcript.

Methods

Sample acquisition

The specimen used for genome sequencing, one female of *Sepioteuthis* sp. 2, the "shiro" type (specimen ID VIEC0000008, ToLID xcSepLess1), was collected using jigging off Tosa Bay in Kochi Prefecture on 2021-04-23. The specimen was shipped frozen at –20 °C to Hiroshima University. The muscle mantle of this specimen was preserved in a 2mL Eppendorf tube and kept at –20 °C.

The specimen used for Hi-C and RNA sequencing (specimen ID VIEC0000009, ToLID xcSepLess2) was an adult specimen collected from Shimane Prefecturen, Hamada, Japan on 2021-05-21 by jigging. The specimens were collected and identified by Gustavo Sanchez (Hiroshima University).

Table 2. Genome assembly data for *Sepioteuthis lessoniana*.

Genome assembly		
Assembly name	xcSepLess1.1	
Assembly accession	GCA_963585895.1	
Alternate haplotype accession	*GCA_963585875.1*	
Assembly level for primary assembly	chromosome	
Span (Mb)	5,056.23	
Number of contigs	13,916	
Number of scaffolds	6,819	
Longest scaffold (Mb)	165.03	
Assembly metric	**Measure**	***Benchmark***
Contig N50 length	0.66 Mb	*≥ 1 Mb*
Scaffold N50 length	96.87 Mb	*= chromosome N50*
Consensus quality (QV)	Primary: 51.5; alternate: 52.7; combined 52.0	*≥ 40*
k-mer completeness	Primary: 84.36%; alternate: 81.77%; combined: 94.54%	*≥ 95%*
BUSCO*	C:73.7%[S:72.7%,D:1.0%],F:4.8%,M:21.4%,n:5,295	*S > 90%* *D < 5%*
Percentage of assembly assigned to chromosomes	86.38%	*≥ 90%*
Organelles	Mitochondrial genome: 16.64 kb	*complete single alleles*
Genome annotation of assembly GCA_963585895.1 at Ensembl		
Number of protein-coding genes	28,970	
Number of non-coding genes	21,583	
Number of gene transcripts	68,971	

* BUSCO scores based on the mollusca_odb10 BUSCO set using version 5.5.0. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023). The xcSepLess1 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the muscle tissue was cryogenically disrupted using the Covaris cryoPREP® Automated Dry Pulverizer (Narváez-Gómez *et al.*, 2023). HMW DNA was extracted using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *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 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 gill tissue of xcSepLess2 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.

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers' instructions. DNA sequencing was performed by the Scientific

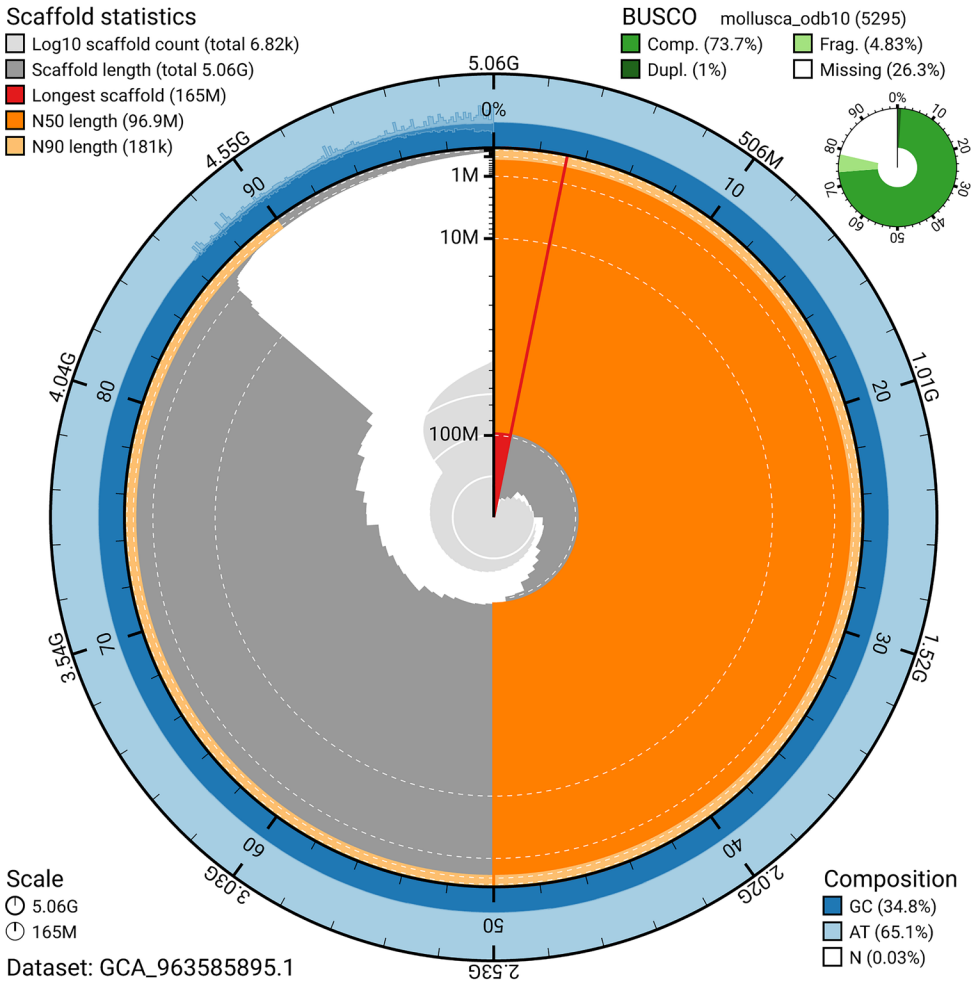

Figure 2. Genome assembly of *Sepioteuthis lessoniana*, xcSepLess1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the 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 mollusca_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963585895.1/dataset/GCA_963585895.1/snail.

Operations core at the WSI on Pacific Biosciences Sequel IIe. Tissue from the somatic tissue of the xcSepLess2 sample was processed for Hi-C sequencing at the WSI Scientific Operations core, using the Arima-HiC v2 kit.and sequenced on the Illumina NovaSeq 6000 instrument. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit, following the manufacturer's instructions. RNA sequencing was performed on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

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 were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

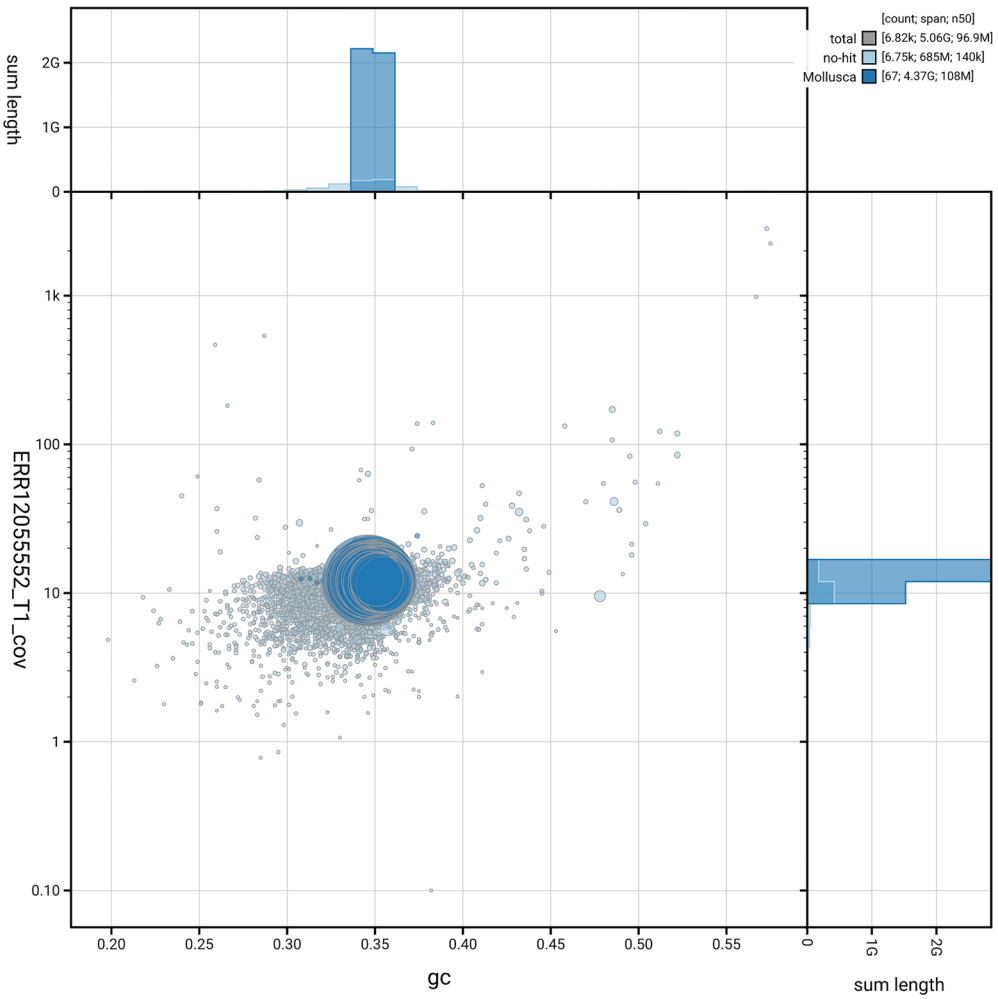

Figure 3. Genome assembly of *Sepioteuthis lessoniana*, xcSepLess1.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. 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 at https://blobtoolkit.genomehubs.org/view/GCA_963585895.1/dataset/GCA_963585895.1/blob.

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. Flat files and maps used in curation were generated via the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was conducted primarily in PretextView (Harry, 2022) and HiGlass (Kerpedjiev *et al.*, 2018), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were amended, and duplicate sequences were tagged and removed. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation.

Assembly quality assessment

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

A Hi-C contact map was produced for the final version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BED-

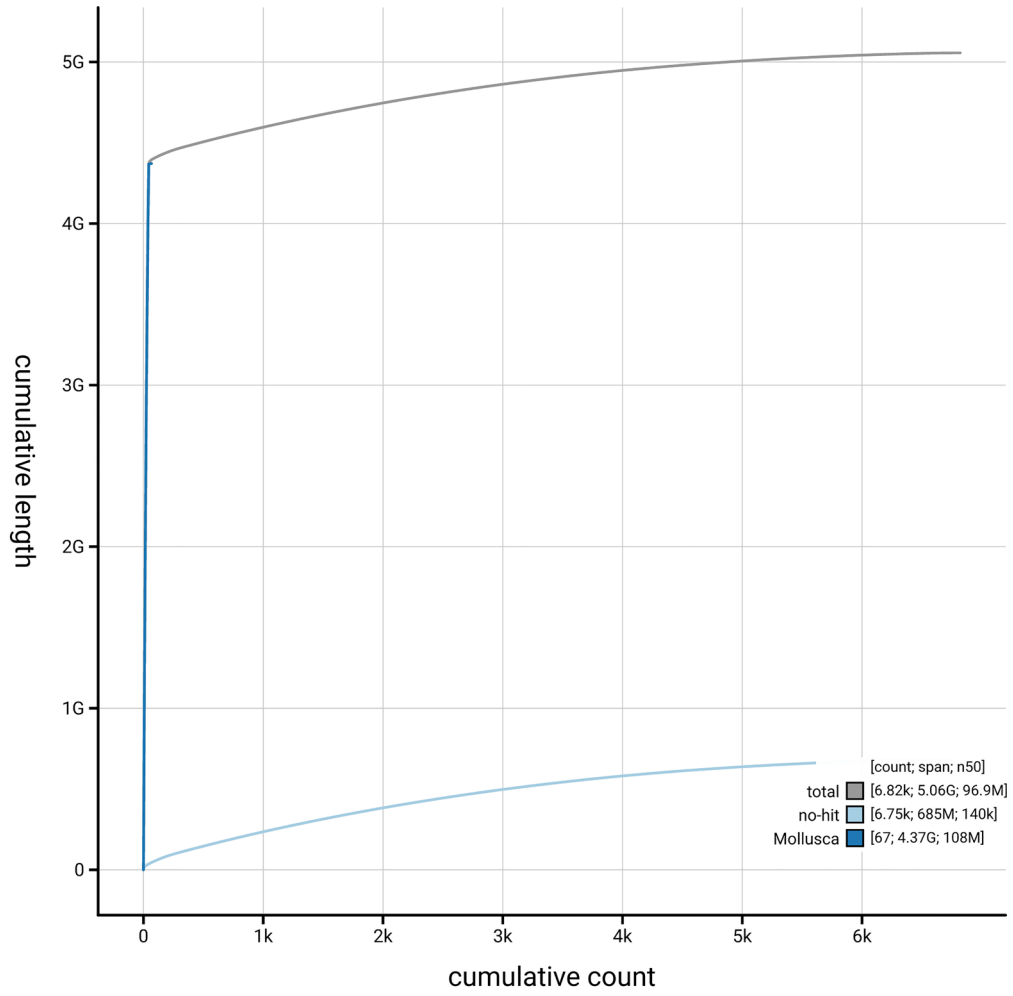

Figure 4. Genome assembly of *Sepioteuthis lessoniana*, xcSepLess1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963585895.1/dataset/GCA_963585895.1/cumulative.

Tools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The blobtoolkit pipeline is a Nextflow port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in 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 lineages, the pipeline aligns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these 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), relying on the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 5 contains a list of relevant software tool versions and sources.

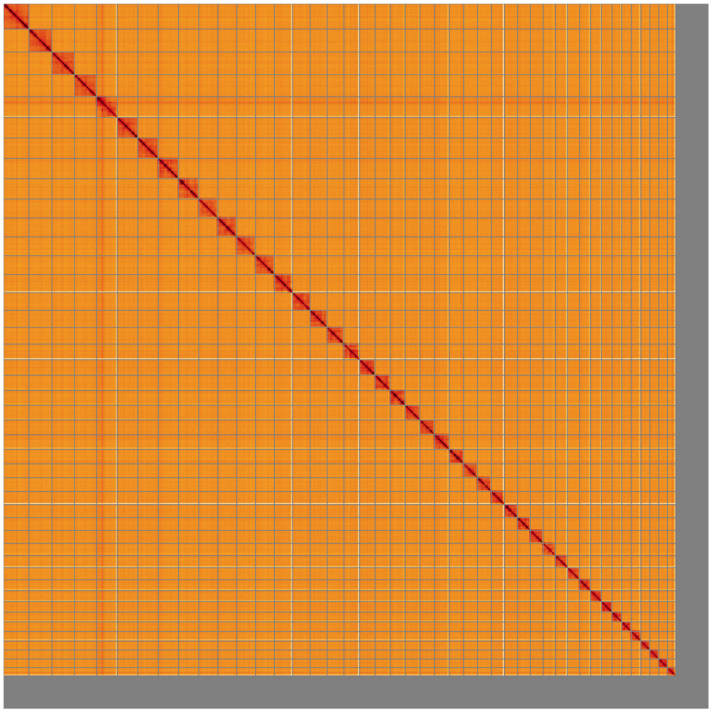

Figure 5. Genome assembly of *Sepioteuthis lessoniana*, xcSepLess1.1: Hi-C contact map of the xcSepLess1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=RfIiARSCSauoz1ZtvLxtng.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Sepioteuthis lessoniana*, xcSepLess1.

INSDC accession	Name	Length (Mb)	GC%
OY757904.1	1	165.03	34.5
OY757905.1	2	148.29	34.5
OY757906.1	3	146.83	34.5
OY757907.1	4	143.61	34.5
OY757908.1	5	135.24	35
OY757909.1	6	133.26	35
OY757910.1	7	133.09	35
OY757911.1	8	131.98	34.5
OY757912.1	9	131.16	34.5
OY757913.1	10	124.05	34.5
OY757914.1	11	122.65	35
OY757915.1	12	122.42	35
OY757916.1	13	122.24	34.5
OY757917.1	14	117.8	35
OY757918.1	15	114.22	35
OY757919.1	16	111.07	35
OY757920.1	17	108.32	34.5
OY757921.1	18	101.85	35
OY757922.1	19	101.23	35.5
OY757923.1	20	97.57	35
OY757924.1	21	96.87	35
OY757925.1	22	96.03	35
OY757926.1	23	95.44	35.5
OY757927.1	24	95.42	35
OY757928.1	25	93.84	35
OY757929.1	26	90.61	34.5
OY757930.1	27	89.63	35
OY757931.1	28	84.68	35.5
OY757932.1	29	84.74	35.5
OY757933.1	30	83.72	35
OY757934.1	31	82.93	35
OY757935.1	32	80.29	35

INSDC accession	Name	Length (Mb)	GC%
OY757936.1	33	78.34	35
OY757937.1	34	78.08	35
OY757938.1	35	71.07	35
OY757939.1	36	70.69	35.5
OY757940.1	37	68.51	35.5
OY757941.1	38	63.29	35.5
OY757942.1	39	62.91	35.5
OY757943.1	40	62.5	35.5
OY757944.1	41	58.21	35.5
OY757945.1	42	57.4	35
OY757946.1	43	57.12	35.5
OY757947.1	44	53.38	35
OY757948.1	MT	0.02	29

Genome annotation

The Ensembl Genebuild annotation system (Aken *et al.*, 2016) was used to generate annotation for the *Sepioteuthis lessoniana* assembly (GCA_963585895.1) in Ensembl Rapid Release at the EBI. Annotation was created primarily through alignment of transcriptomic data to the genome, with gap filling via protein-to-genome alignments of a select set of proteins from UniProt (UniProt Consortium, 2019).

Metagenome assembly

The metagenome assembly was generated using MetaMDBG (Benoit *et al.*, 2024) and binned using MetaBAT2 (Kang *et al.*, 2019), MaxBin (Wu *et al.*, 2014), bin3C (DeMaere & Darling, 2019), and MetaTOR. The resulting bin sets of each binning algorithm were optimised and refined using MAGScoT (Rühlemann *et al.*, 2022). PROKKA (Seemann, 2014) was used to identify tRNAs and rRNAs in each bin, CheckM (Parks *et al.*, 2015) (checkM_DB release 2015-01-16) was used to assess bin completeness/contamination, and GTDB-TK (Chaumeil *et al.*, 2022) (GTDB release 214) was used to taxonomically classify bins. Taxonomic replicate bins were identified

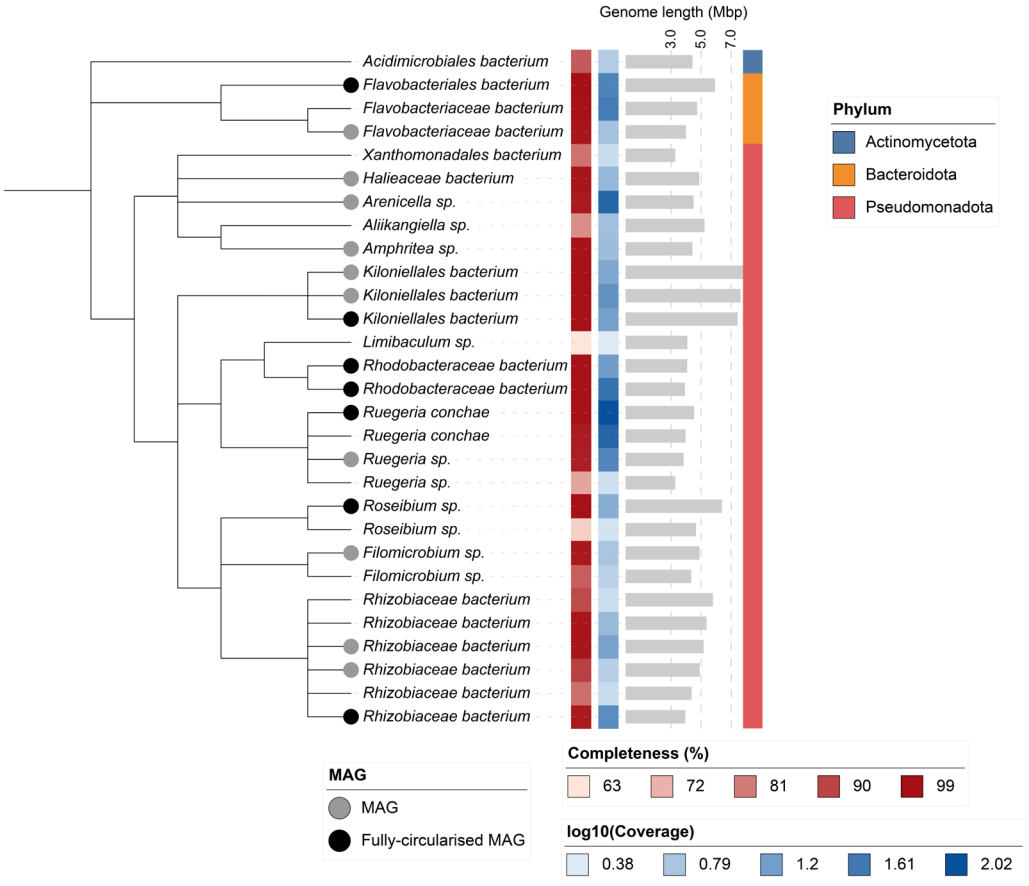

Figure 6. Cladogram showing the taxonomic placement of metagenome bins, constructed using NCBI taxonomic identifiers with *taxonomizr* and annotated in iTOL. Colours indicate phylum-level taxonomy. Additional tracks show sequencing coverage (log_{10}), genome size (Mbp), and completeness. Bins that meet the criteria for MAGs are marked with a grey circle; fully circularised MAGs are marked in black.

Table 4. Quality metrics and taxonomic assignments of the binned metagenomes.

NCBI taxon	Taxid	GTDB taxonomy	Quality	Size (bp)	Contigs	Circular	Mean coverage	Completeness (%)	Contamination (%)
Ruegeria sp.	259304	g__Ruegeria	High	3,805,421	1	No	29.78	97.34	0.53
Rhodobacteraceae bacterium	157276	g__SZUA-547	High	3,886,458	1	Yes	46.65	99.6	0.6
Rhizobiaceae bacterium	271151	f__Rhizobiaceae	High	3,917,026	1	Yes	25.66	97.91	1.23
Ruegeria conchae	981384	s__Ruegeria conchae	High	3,942,089	1	Yes	66.83	97.66	0.32
Flavobacteriaceae bacterium	165436	g__SMXJ01	High	3,961,667	3	No	6.69	98.76	0.41
Rhodobacteraceae bacterium	157276	g__SZUA-547	High	4,052,132	1	Yes	16.84	99.6	0.6
Amphritea sp.	981605	g__Amphritea	High	4,385,521	2	No	7.67	99.29	1
Arenicella sp.	1586337	g__Arenicella	High	4,474,314	19	Partial	65.39	98.17	0.61
Ruegeria conchae	981384	s__Ruegeria conchae	High	4,502,159	2	Yes	104.23	99.49	0.43
Halieaceae bacterium	2735679	f__Halieaceae	High	4,838,237	1	No	8.96	98.7	1.99
Filomicrobium sp.	293328	g__Filomicrobium	High	4,869,283	35	Partial	6.22	98.39	2.35
Rhizobiaceae bacterium	271151	f__Rhizobiaceae	High	4,883,131	32	No	4.87	91.73	2.09
Rhizobiaceae bacterium	271151	f__Rhizobiaceae	High	5,142,339	7	No	14.59	98.8	4.78
Flavobacteriales bacterium	213322	o__Flavobacteriales	High	5,895,559	1	Yes	29.94	99.73	1.34
Roseibium sp.	1936171	g__Roseibium	High	6,353,716	1	Yes	10.96	99.58	0.63
Kiloniellales bacterium	1667039	g__JACOMY01	High	7,402,078	1	Yes	15.46	99.57	0.87
Kiloniellales bacterium	1667039	g__JACOMY01	High	7,596,607	1	No	21.94	99.57	0.43
Kiloniellales bacterium	1667039	g__JACOMY01	High	7,789,244	1	No	13.27	99.57	0.22
Xanthomonadales bacterium	412058	f__SZUA-36	Medium	3,248,922	30	No	3.55	83.14	0.91
Ruegeria sp.	259304	s__Ruegeria sp003443535	Medium	3,249,342	27	No	3.29	74.63	0
Limibaculum sp.	2036014	g__Limibaculum	Medium	4,064,720	170	No	2.4	63.37	1.2
Filomicrobium sp.	293328	g__Filomicrobium	Medium	4,321,939	42	No	4.53	86.38	1.81
Rhizobiaceae bacterium	271151	g__JAALLB01	Medium	4,339,548	49	No	3.58	83.5	1.34
Acidimicrobiales bacterium	310071	f__Bin134	Medium	4,404,047	14	No	5.15	87.18	2.99
Roseibium sp.	1936171	g__Roseibium	Medium	4,625,544	74	No	2.97	67.24	3.45
Flavobacteriaceae bacterium	165436	g__GCA-2733415	Medium	4,710,532	7	No	36.9	99.53	7.38
Aliikangiella sp.	1920244	g__Aliikangiella	Medium	5,204,448	13	No	7.08	78.38	0.27
Rhizobiaceae bacterium	271151	f__Rhizobiaceae	Medium	5,326,527	32	No	8.06	98.8	5.26
Rhizobiaceae bacterium	271151	g__SPNT01	Medium	5,757,256	129	Partial	3.41	89.74	8.83

Table 5. Software tools: versions and sources.

Software tool	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
bin3C	0.3.3	https://github.com/cerebis/bin3C
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
CheckM	1.2.1	https://github.com/Ecogenomics/CheckM
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
dRep	3.4.0	https://github.com/MrOlm/drep
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
GTDB-TK	2.3.2	https://github.com/Ecogenomics/GTDBTk
Hifiasm	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	44086069ee7d4d3f6f3f0012569789ec138f42b84aa44357826c0b6753eb28de	https://github.com/higlass/higlass
MaxBin	2.7	https://sourceforge.net/projects/maxbin/
MerquryFK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
MetaBat2	2.15-15-gd6ea400	https://bitbucket.org/berkeleylab/metabat/src/master/
MetaTOR	-	https://github.com/koszullab/metaTOR
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.04.1	https://github.com/nextflow-io/nextflow
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
PROKKA	1.14.5	https://github.com/vdejager/prokka
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.4.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.1a.2	https://github.com/c-zhou/yahs

using dRep (Olm *et al.*, 2017) with default settings (95% ANI threshold). The final bin set was filtered for bacteria and archaea. All bins were assessed for quality and categorised as metagenome-assembled genomes (MAGs) if they met the following criteria: contamination ≤ 5%, presence of 5S, 16S, and 23S rRNA genes, at least 18 unique tRNAs, and either ≥ 90% completeness or ≥ 50% completeness with fully circularised chromosomes. Bins that did not meet these thresholds, or were identified as taxonomic replicates of MAGs, were retained as 'binned metagenomes' provided they had ≥ 50% completeness and ≤ 10% contamination. A cladogram based on NCBI taxonomic assignments was generated using the 'taxonomizr' package in R. The tree was visualised and annotated using iTOL (Letunic & Bork, 2024). Software tool versions and sources are given in Table 5.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. 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 undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute) and in some circumstances other Tree of Life collaborators.

Data availability

European Nucleotide Archive: Sepioteuthis lessoniana (bigfin reef squid). Accession number PRJEB64979; https://identifiers.org/ena.embl/PRJEB64979. The genome sequence is released openly for reuse. The *Sepioteuthis lessoniana* genome sequencing initiative is part of the Aquatic Symbiosis Genomics (ASG) project (https://www.ebi.ac.uk/ena/browser/view/PRJEB43743). All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in Table 1 and Table 2.

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.13960213.

Members of the Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.14870789.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.15495178.

Members of the European Bioinformatics Institute ASG Data Portal team are listed here: https://doi.org/10.5281/zenodo.10076466.

Members of the Wellcome Sanger Institute/Aquatic Symbiosis Genomics Project Leadership are listed here: https://doi.org/10.5281/zenodo.10184833.

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

Aken BL, Ayling S, Barrell D, *et al.*: **The ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
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

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

Arnold JM: **Embryonic development of the squid.** In: Gilbert, D. L., Adelman, W. J., and Arnold, J. M. (eds.) *Squid as experimental animals.* Boston, MA: Springer US, 1990; 77–90.
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

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for LI PacBio.** *protocols.io.* 2023.
Publisher Full Text

Benoit G, Raguideau S, James R, *et al.*: **High-quality metagenome assembly from long accurate reads with metaMDBG.** *Nat Biotechnol.* 2024; **42**(9): 1378–1383.
PubMed Abstract | Publisher Full Text | Free 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

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

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality**

assessment of genome assemblies. *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Chaumeil PA, Mussig AJ, Hugenholtz P, *et al.*: **GTDB-Tk v2: memory friendly classification with the genome taxonomy database.** *Bioinformatics.* 2022; **38**(23): 5315–5316.
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**

Cheng SH, Anderson FE, Bergman A, *et al.*: **Molecular evidence for co-occurring cryptic lineages within the *Sepioteuthis cf. lessoniana* species complex in the Indian and Indo-West Pacific Oceans.** *Hydrobiologia.* 2014; **725**(1): 165–188.
Publisher 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**

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**

DeMaere MZ, Darling AE: **bin3C: exploiting Hi-C sequencing data to accurately resolve metagenome-assembled genomes.** *Genome Biol.* 2019; **20**(1): 46.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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

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**

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps**. 2022.
Reference Source

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**

Imai H, Aoki M: **Genetic diversity and genetic heterogeneity of bigfin reef squid "*Sepioteuthis lessoniana*" species complex in Northwestern Pacific Ocean.** In: Caliskan, M. (ed.) *Analysis of genetic variation in animal.* InTech, 2012.
Publisher 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

Kang DD, Li F, Kirton E, *et al.*: **MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies.** *PeerJ.* 2019; **7**: e7359.
PubMed Abstract | **Publisher Full Text** | **Free 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**

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**

Letunic I, Bork P: **Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool.** *Nucleic Acids Res.* 2024; **52**(W1): W78–W82.
PubMed Abstract | **Publisher Full Text** | **Free 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.* 2021; **38**(10): 4647–4654.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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

Nakajima R, Lajbner Z, Kuba MJ, *et al.*: **Squid adjust their body color according to substrate.** *Sci Rep.* 2022; **12**(1): 5227.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Narváez-Gómez JP, Mbye H, Oatley G, *et al.*: **Sanger Tree of Life sample homogenisation: covaris cryoPREP® automated dry pulverizer V.1.** *protocols.io.* 2023.
Publisher Full Text

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023.
Publisher Full Text

Olm MR, Brown CT, Brooks B, *et al.*: **dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication.** *ISME J.* 2017; **11**(12): 2864–2868.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Parks DH, Imelfort M, Skennerton CT, *et al.*: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome Res.* 2015; **25**(7): 1043–55.
PubMed Abstract | **Publisher Full Text** | **Free 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**

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, 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**

Rühlemann MC, Wacker EM, Ellinghaus D, *et al.*: **MAGScoT: a fast, lightweight and accurate bin-refinement tool.** *Bioinformatics.* 2022; **38**(24): 5430–5433.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Sakurai Y, Ikeda Y: **Effect of visual lateralization on the spatial position of individuals within a school of oval squid (*Sepioteuthis lessoniana*).** *J Comp Physiol A Neuroethol Sens Neural Behav Physiol.* 2024; **210**(3): 381–398.
PubMed Abstract | **Publisher Full Text**

Seemann T: **Prokka: rapid prokaryotic genome annotation.** *Bioinformatics.* 2014; **30**(14): 2068–2069.
PubMed Abstract | **Publisher Full Text**

Shigeno S, Tsuchiya K, Segawa S: **Conserved topological patterns and heterochronies in loliginid cephalopods: comparative developmental morphology of the oval squid *Sepioteuthis lessoniana*.** *Invertebr Reprod Dev.* 2001; **39**(3): 161–174.
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

Sugimoto C, Ikeda Y: **Ontogeny of schooling behavior in the oval squid *Sepioteuthis lessoniana*.** *Fish Sci.* 2012; **78**(2): 287–294.
Publisher Full Text

Sugimoto C, Yanagisawa R, Nakajima R, *et al.*: **Observations of schooling behaviour in the oval squid *Sepioteuthis lessoniana* in coastal waters of Okinawa Island.** *Mar Biodivers Rec.* 2013; **6**: e34.
Publisher Full Text

Tomano S, Sanchez G, Kawai K, *et al.*: **Contribution of *Sepioteuthis* sp. 1 and *Sepioteuthis* sp. 2 to oval squid fishery stocks in western Japan.** *Fish Sci.* 2016; **82**(4): 585–596.
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**

UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
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

Wu YW, Tang YH, Tringe SG, *et al.*: **MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm.** *Microbiome.* 2014; **2**(1): 26.
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 09 January 2026

https://doi.org/10.21956/wellcomeopenres.27060.r142580

?

Ole Kristian Tørresen
Department of Biosciences, University of Oslo CEES (Ringgold ID: 206081), Oslo, Oslo, Norway

The authors have sequenced and assembled the genome of bigfin reef squid, a squid occurring in the Indo-Pacific Ocean.
The data note follows the usual setup, and is in general concise and to the point for most information, but with some details lacking. The biggest issue I have with this data note is the metagenomic part.

What is the reason for doing this particular analysis? Is it known that these squids have lots of different bacteria in their tissues? (Which seems quite strange.)

On which sample (tissue) was this performed? This is not stated as far as I can see. Both xcSepLess1 and xcSepLess2 are from somatic tissue, and as far as I understand, should not contain a lot of bacteria.

As it is now, the metagenomic part of the data note seems disconnected from the rest, with no justification and lacking in detail around the source of the sample.

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: Genome assembly

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

Reviewer Report 07 January 2026

https://doi.org/10.21956/wellcomeopenres.27060.r142577

David EK Ferrier
University of St Andrews, Scotland, UK

The rationale is clear. Besides the need for sequencing this mollusc as part of the Earth Biogenome Project and the Aquatic Symbiosis Initiative, there is the added interest of this animal being amenable to lab culture and hence experimental approaches for a variety of biological questions as well as aquaculture (as summarized briefly in the Introduction).

The various protocols are certainly appropriate and the work is of a high technical standard.

The materials and methods descriptions meet accepted standards for these data notes, and the datasets such as the assemblies of the two haplotypes are accessible and useable.

There are a couple of minor typos for correction.
1) Typo in Background, third line: insert ‘to’ to read, “that belongs to the”.

2) Genome sequence report section, first paragraph, line 8: delete an unnecessary hyphen to read, “generate”.

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.

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 02 January 2026

https://doi.org/10.21956/wellcomeopenres.27060.r142579

Erich M. Schwarz
Cornell University, Ithaca, New York, USA

The genome assemblies of Sanchez et al. are well done and well described. I have only a few questions.

1. What is the correct full name for the species being analyzed? In the title, the authors refer to it as "*Sepioteuthis lessoniana* d'Orbigny, 1826"; but at the start of the Background, the authors refer to it as "*Sepioteuthis lessoniana Férussac*, 1831". Which of these full species names is correct, and could the authors please stick to the correct one?

2. In the Sample Acquisition section of the Methods, the authors give a species identification for the specimen used in genome sequencing ("one female of Sepioteuthis sp. 2"), but do not give a species identification for the specimen used in Hi-C and RNA-sequencing ("an adult specimen"). Could the authors provide a species identification for the second specimen?

3. Would it be possible to provide a scale bar for Figure 1?

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: Nematode genome assembly and analysis

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.