Check for updates

DATA NOTE

ERGA-BGE genome of *Stigmatoteuthis arcturi* Robson, 1948: the jewelled squid

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

Fernando Ángel Fernández-Álvarez[1,2], Ainhoa Bernal-Bajo[3], Nuria Escudero[4], María Conejero[5], Ana Riesgo[5], Rosa Fernández[4], Rita Monteiro[6], Astrid Böhne[6], Laura Aguilera[7,8], Marta Gut[7,8], Tyler S Alioto[7,8], Francisco Câmara Ferreira[7,8], Fernando Cruz[7,8], Jèssica Gómez-Garrido[7,8], Leanne Haggerty[9], Fergal Martin[9], Tom Brown[10,11]

[1]Institut de Ciencies del Mar (ICM-CSIC), CSIC, Barcelona, Spain
[2]Instituto Español de Oceanografía, Centro Oceanográfico de Gijón (IEO, CSIC), Av. del Príncipe de Asturias, 70 Bis, Gijón, 33212, Spain
[3]Centro Oceanográfico de Baleares, Instituto Español de Oceanografía (IEO-CSIC), Palma, Spain
[4]Institut de Biologia Evolutiva (UPF-CSIC), Barcelona, Spain
[5]Museo Nacional de Ciencias Naturales- CSIC, Madrid, Spain
[6]Leibniz Institute for the Analysis of Biodiversity Change - Museum Koenig Bonn, Bonn, Germany
[7]Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain
[8]Universitat de Barcelona (UB), Barcelona, Spain
[9]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK
[10]Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany
[11]Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Berlin, Germany

v1 **First published:** 04 Sep 2025, **5**:270
https://doi.org/10.12688/openreseurope.20439.1
Latest published: 04 Sep 2025, **5**:270
https://doi.org/10.12688/openreseurope.20439.1

Abstract

The *Stigmatoteuthis arcturi* reference genome offers a valuable resource for understanding the evolutionary patterns of oceanic squids, who perform important ecological roles as both predators and prey in mesopelagic and deep-sea environments, while their genomes remain understudied. The entirety of the genome sequence was assembled into 46 contiguous chromosomal pseudomolecules. This chromosome-level assembly encompasses 3.25 Gb, composed of 1,268 contigs and 497 scaffolds, with contig and scaffold N50 values of 11.0 Mb and 74.6 Mb, respectively.

Keywords

Stigmatoteuthis arcturi, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, jewelled squid

Open Peer Review

Approval Status ✓ ✓ ?

	1	2	3
version 1 04 Sep 2025	✓ view	✓ view	? view

1. **Masa-Aki Yoshida**, Shimane University, Shiman, Japan
2. **José M Martín-Durán**, Queen Mary University of London, London, UK
3. **Marco Gerdol**, University of Trieste, Trieste, Italy

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

This article is included in the Horizon Europe gateway.

This article is included in the Genome Reports from the Biodiversity Genomics Europe Project collection.

Corresponding author: Tom Brown (brown@izw-berlin.de)

Author roles: Fernández-Álvarez FÁ: Investigation, Methodology, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Bernal-Bajo A**: Investigation, Resources, Writing – Review & Editing; **Escudero N**: Methodology, Supervision, Validation, Writing – Review & Editing; **Conejero M**: Investigation, Resources, Writing – Review & Editing; **Riesgo A**: Investigation, Resources, Writing – Review & Editing; **Fernández R**: Funding Acquisition, Methodology, Supervision, Validation, Writing – Review & Editing; **Monteiro R**: Methodology, Supervision, Validation, Writing – Review & Editing; **Böhne A**: Funding Acquisition, Methodology, Supervision, Validation, Writing – Review & Editing; **Aguilera L**: Investigation, Writing – Review & Editing; **Gut M**: Funding Acquisition, Investigation, Supervision, Writing – Review & Editing; **Alioto TS**: Data Curation, Formal Analysis, Funding Acquisition, Supervision, Writing – Review & Editing; **Câmara Ferreira F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Cruz F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Gómez-Garrido J**: Data Curation, Formal Analysis, Writing – Review & Editing; **Haggerty L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Martin F**: Data Curation, Formal Analysis, Funding Acquisition, Supervision, Writing – Review & Editing; **Brown T**: Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101059492 (Biodiversity Genomics Europe [BGE]). This project received funding from Horizon Europe under the Biodiversity, Circular Economy and Environment (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. Funding was provided by the Spanish Ministry of Science, Innovation and Universities (OCTOSET, Ref. RTI2018-097908-B-I00; ECOPHYN, Ref. PID2021-126824NB-C32; MCIU/AEI/FEDER, EU), and the Spanish government through the "Severo Ochoa Center of Excellence" accreditation (CEX2019-000928-S). F.Á.F.-Á. was supported by a Beatriu de Pinós fellowship from Secretaria d'Universitats i Recerca del Departament de Recerca i Universitats of the Generalitat de Catalunya (Ref. BP 2021 00035) and a Ramón y Cajal fellowship (Ref. RYC2023-043494-I) funded by MCIN/AEI /10.13039/501100011033 and FSE+.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Fernández-Álvarez FÁ *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: Fernández-Álvarez FÁ, Bernal-Bajo A, Escudero N *et al.* **ERGA-BGE genome of *Stigmatoteuthis arcturi* Robson, 1948: the jewelled squid [version 1; peer review: 2 approved, 1 approved with reservations]** Open Research Europe 2025, **5**:270 https://doi.org/10.12688/openreseurope.20439.1

First published: 04 Sep 2025, **5**:270 https://doi.org/10.12688/openreseurope.20439.1

Introduction

Stigmatoteuthis arcturi Robson, 1948 belongs to the family Histioteuthidae Verrill, 1880–1881, known as the jewelled squids, which are one of the most important components of the diet of endangered megafauna, such as sperm whales (Clarke, 2006). Jewelled squids are characterised by a unique morphology with a number of photophores on their skin to break its shadow and deceive predators from deeper waters. They also possess high levels of asymmetry in their bodies, peaking in a large difference in size, morphology and pigmentation of their eyes, themselves specialised to different tasks (Thomas *et al.*, 2017). While the larger left eye looks at the dim light coming from the surface to spot their megafaunal predators, the smaller right eye looks towards the bottom looking for bioluminescence of their micronekton prey. *S. arcturi* is one of the three allopatric species of the genus *Stigmatoteuthis* Pfeffer, 1900, which are characterised by duplicated terminal parts of the male reproductive system and with subtle morphological differences among them, which can only be recognised in mature males (Young & Vecchione, 2016). It distributes in tropical and subtropical Atlantic offshore mesopelagic waters and as with any other cephalopod, *S. arcturi* is fast growing, fueled by a very intense predatory activity. Jewelled squids are paratenic hosts of parasitic ascarid helminths, such as *Anysakis* Dujardin, 1845 and other nematodes (Palomba *et al.*, 2021). They transfer these parasites to higher trophic level hosts, such as commercially important swordfishes and endangered toothed whales, where these parasites finish their life cycles.

Jewelled squids are the largest component of sperm whales diets in many areas (Clarke, 2006) and an important prey item for many other megafaunal species, such as dolphins, sharks and swordfishes. Thus, they represent an important energy and biomass link among lower and upper trophic levels in pelagic oceanic food webs. Jewelled squids have an unsettled position in the phylogeny of oceanic squids (Fernández-Álvarez *et al.*, 2022). The genome of *S. arcturi* is a valuable tool to solve the phylogenetic position of jewelled squids within oceanic squids.

The generation of this reference resource was coordinated by the European Reference Genome Atlas (ERGA) initiative's Biodiversity Genomics Europe (BGE) project, supporting ERGA's aims of promoting transnational cooperation to promote advances in the application of genomics technologies to protect and restore biodiversity (Mazzoni *et al.*, 2023).

Materials & methods

ERGA's sequencing strategy includes Oxford Nanopore Technology (ONT) and/or Pacific Biosciences (PacBio) for long-read sequencing, along with Hi-C sequencing for chromosomal architecture, Illumina Paired-End (PE) for polishing (i.e. recommended for ONT-only assemblies), and RNA sequencing for transcriptomic profiling, to facilitate genome assembly and annotation.

Sample and sampling information

Ainhoa Bernal-Bajo sampled one specimen of *Stigmatoteuthis arcturi*, which was identified based on morphology by Fernando Ángel Fernández-Álvarez, from international Atlantic waters near Azores and the Canary Islands on 28th February 2023. As sampling was performed in international waters, no permit is required. As the squid is not considered an experimental animal according to the Spanish directive RD 53/2013, there is no need for an approval by an ethical committee. Sampling was performed using a Mesopelagos net.

The specimen was euthanized by freezing at -80 °C. Until DNA and RNA extractions, samples were preserved at -80 °C.

Vouchering information

Physical reference material for the here sequenced specimen has been deposited in the National Museum of Natural History of Madrid (MNCN-CSIC) https://www.mncn.csic.es/es/colecciones under the accession number MNCN 15.06/521.

Frozen reference tissue material of skin is available from the same individual at the Biobank National Museum of Natural History of Madrid (MNCN-CSIC) https://www.mncn.csic.es/es/colecciones under the ID MNCN-ADN-151723.

An electronic voucher image of the sequenced individual is available from ERGA's EBI BioImageArchive dataset https://www.ebi.ac.uk/biostudies/bioimages/studies/S-BIAD1012?query=ERGA under accession IDs SAMEA114541340_1.jpg, SAMEA114541340_2.jpg, and SAMEA114541340_3.jpg.

Genetic information

The estimated genome size, based on ancestral taxa, is 3.2 Gb. This is a diploid genome with a haploid number of 46 chromosomes (2n=92). All information for this species was retrieved from Genomes on a Tree (Challis *et al.*, 2023).

DNA/RNA processing

DNA was extracted from arm tissue using the Blood & Cell Culture DNA Midi Kit (Qiagen) following the manufacturer's instructions. DNA quantification was performed using a Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific), and DNA integrity was assessed using a Genomic DNA 165 Kb Kit (Agilent) on the Femto Pulse system (Agilent). The DNA was stored at +4°C until used.

RNA was extracted using an RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA was extracted from four different specimen parts: muscle, skin, arm tissue and tentacle stalk. RNA quantification was performed using

the Qubit RNA BR kit and RNA integrity was assessed using a Bioanalyzer 2100 system (Agilent) RNA 6000 Nano Kit (Agilent). RNA was equimolarly pooled for the library preparation and stored at -80°C until used.

Library preparation and sequencing

For long-read whole genome sequencing, a library was prepared using the SQK-LSK114 Kit (Oxford Nanopore Technologies, ONT) and was sequenced on a PromethION 24 A Series instrument (ONT). A short-read whole genome sequencing library was prepared using the KAPA Hyper Prep Kit (Roche). A Hi-C library was prepared from arm tissue using the Dovetail Omni-C kit (Cantata Bio), followed by the KAPA Hyper Prep kit for Illumina sequencing (Roche). The RNA library from the pooled sample was prepared using the KAPA mRNA Hyper prep kit for Illumina sequencing (Roche). The short-read sequencing libraries were processed on a NovaSeq 6000 instrument (Illumina). In total, 180.5 Gb Oxford Nanopore, 203.5 Gb Illumina WGS shotgun, and 195.3 Gb HiC data were sequenced to generate the assembly.

Genome assembly methods

The genome was assembled using the CNAG CLAWS pipeline (Gomez-Garrido, 2024). Briefly, reads were preprocessed for quality and length using Trim Galore v0.6.7 and Filtlong v0.2.1, and initial contigs were assembled using NextDenovo v2.5.0, followed by polishing of the assembled contigs using HyPo v1.0.3, removal of retained haplotigs using purge-dups v1.2.6 and scaffolding with YaHS v1.2a. Finally, assembled scaffolds were curated via manual inspection using Pretext v0.2.5 with the Rapid Curation Toolkit (https://gitlab.com/wtsi-grit/rapid-curation) to remove any false joins and incorporate any sequences not automatically scaffolded into their respective locations in the chromosomal pseudomolecules (or super-scaffolds). Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (https://doi.org/10.48546/workflowhub.workflow.1103.2).

Genome annotation methods

A gene set was generated using the Ensembl Gene Annotation system (Aken *et al.*, 2016), primarily by aligning publicly available short-read RNA-seq data from BioSample: SAMEA117648700 to the genome. Gaps in the annotation were filled via protein-to-genome alignments of a select set of clade-specific proteins from UniProt ("UniProt: A Worldwide Hub of Protein Knowledge," 2019), which had experimental evidence at the protein or transcript level. At each locus, data were aggregated and consolidated, prioritising models derived from RNA-seq data, resulting in a final set of gene models and associated non-redundant transcript sets. To distinguish true isoforms from fragments, the likelihood of each open reading frame (ORF) was evaluated against known metazoan proteins. Low-quality transcript models, such as those showing evidence of fragmented ORFs, were removed. In cases where RNA-seq data were fragmented or absent, homology data were prioritised, favouring longer transcripts with strong intron support from short-read data. The resulting gene models were classified into two categories: protein-coding, and long non-coding. Models that did not overlap protein-coding genes, and were constructed from transcriptomic data were considered potential lncRNAs. Potential lncRNAs were further filtered to remove single-exon loci due to their unreliability. Putative miRNAs were predicted by performing a BLAST search of miRBase (Kozomara *et al.*, 2019) against the genome, followed by RNAfold analysis (Gruber *et al.*, 2008). Other small non-coding loci were identified by scanning the genome with Rfam (Kalvari *et al.*, 2018) and passing the results through Infernal (Nawrocki & Eddy, 2013). Summary analysis of the released annotation was performed using the ERGA-BGE Genome Report ANNOT Galaxy workflow (https://doi.org/10.48546/workflowhub.workflow.1096.1), incorporating tools such as AGAT v1.2, OMArk v0.3 and BUSCO (v5.5.0).

Results

Genome assembly

The genome assembly has a total length of 3,249,387,216 bp in 497 scaffolds (Figure 1 & Figure 2), with a GC content of 34.92%. The assembly has a contig N50 of 10,980,000 bp and L50 of 88 and a scaffold N50 of 74,550,748 bp and L50 of 19. The assembly has a total of 771 gaps, totaling 154.2 kb in cumulative size. The single-copy gene content analysis using the Metazoa database with BUSCO (Manni *et al.*, 2021) resulted in 94.8% completeness (94.0% single and 0.8% duplicated). 80.5% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 35.4 as calculated by Merqury (Rhie *et al.*, 2020).

Genome annotation

The genome annotation consists of 12,987 protein-coding genes with an associated 21,447 transcripts, in addition to 3,561 non-coding genes (Table 1). Using the longest isoform per transcript, the single-copy gene content analysis using the Metazoa_odb10 database with BUSCO resulted in 93.8% completeness. Using the OMAmer Myomorpha database for OMArk (Nevers *et al.*, 2025) resulted in 93.31% completeness and 74.74% consistency (Table 2).

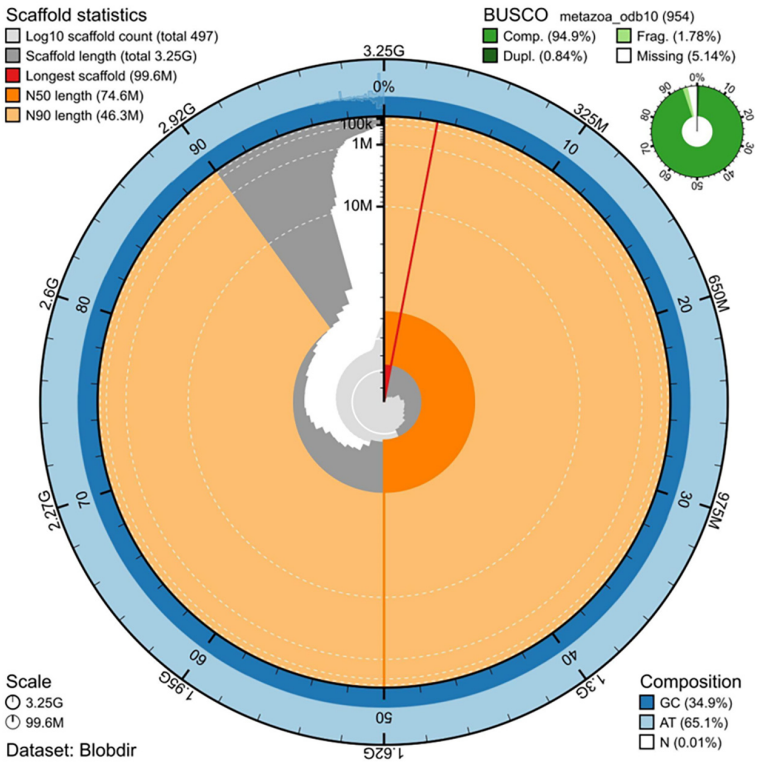

Figure 1. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 3,249,387,216 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (99,556,817 bp, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (74,550,748 and 46,349,533 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log-scale, with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT, and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated, and missing BUSCO genes found in the assembled genome from the Metazoa database (odb10) is shown in the top right.

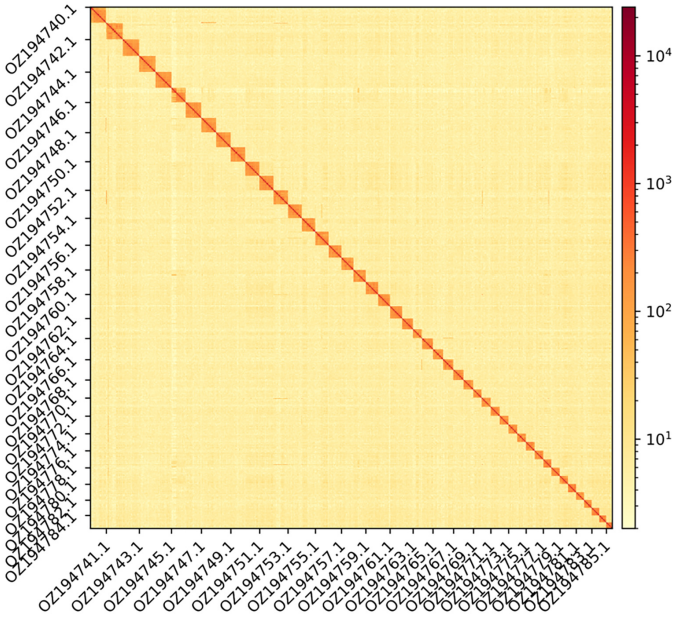

Figure 2. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 25 kb bin size for plotting. Names of alternate chromosomes have been labelled on x- and y-axes for visualisation purposes.

Table 1. **Statistics from assembled gene models.**

	No. genes	No. transcripts	Mean gene length (bp)	No. single-exon genes	Mean exons per transcript
mRNA	12,987	21,447	42,913	519	9.0
snoRNA	219	219	170	219	1
lncRNA	1,300	1,405	13,735	438	2.0
snRNA	837	837	133	837	1
rRNA	1,545	1,545	229	1,545	1
scRNA	5	5	131	5	1
tRNA	12,259	12,259	77	12,259	1

Table 2. **Annotation completeness and consistency scores calculated by BUSCO run in protein mode (metazoa_odb10) and OMArk (Lophotrochozoa).**

	Complete	Singular	Duplicated	Fragmented	Missing
BUSCO	895 (93.8%)	887 (93.0%)	8 (0.8%)	21 (2.2%)	38 (4.0%)
OMArk	2,009 (93.31%)	1,890 (87.78%)	119 (5.53%)	-	144 (6.69%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	9,706 (74.74%)	1,163 (8.96%)	0 (0.0%)	2,118 (16.31%)	

Data availability

S. arcturi and the related genomic study were assigned to Tree of Life ID (ToLID) 'xcStiArct1' and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB81320. The sample information is available at the following BioSample accessions: SAMEA114541340, SAMEA114541342, SAMEA114541343, and SAMEA114541354. The genome assembly is accessible from ENA under accession number GCA_964276865.1. Sequencing data produced as part of this project are available from ENA at the following accessions: ERX13202607, ERX13202608, ERX13202609 and ERX13202610. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) document available at https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Stigmatoteuthis_arcturi/xcStiArct1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/2053936.

Author contributions

AB-B collected the samples, FÁF-Á identified the species, FÁF-Á sampled and preserved biological material and provided metadata, MC and AR provided material and information for vouchering and barcoding, AsB provided sampling and metadata support and management, RM, NE, RF, and AsB provided support in sampling, shipping of biological material, metadata collection, and management, LA and MG extracted DNA, prepared libraries, and performed sequencing, FCF, FC and JGG performed genome assembly and curation under the supervision of TSA, LH, and FM performed genome annotation, TB generated the analysis and report. All authors contributed to the writing, review, and editing of this genome note and read and approved the final version.

Acknowledgements

The specimen was captured during the oceanic research cruise DESAFIO (Ref. PID2020-118118RB-I00). We would like to acknowledge the assembly reviewer, Adama Ndar, from Genoscope. The authors acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI Freiburg.

References

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**

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**

Clarke M: **Oceanic cephalopod distribution and species diversity in the eastern north Atlantic.** *Arquipelago.* 2006; **23A**: 27–46.
Reference Source

Fernández-Álvarez FÁ, Taite M, Vecchione M, *et al.*: **A phylogenomic look into the systematics of oceanic squids (order Oegopsida).** *Zool J Linn Soc.* 2022; **194**(4): 1212–1235.
Publisher Full Text

Gomez-Garrido J: **CLAWS (CNAG's long-read assembly workflow in Snakemake).** WorkflowHub, 2024.
Publisher Full Text

Gruber AR, Lorenz R, Bernhart SH, *et al.*: **The Vienna RNA websuite.** *Nucleic Acids Res.* 2008; **36**(Web Server issue): W70–W74.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kalvari I, Nawrocki EP, Argasinska J, *et al.*: **Non-coding RNA analysis using the Rfam database.** *Curr Protoc Bioinformatics.* 2018; **62**(1): e51.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kozomara A, Birgaoanu M, Griffiths-Jones S: **miRBase: from microRNA sequences to function.** *Nucleic Acids Res.* 2019; **47**(D1): D155–D162.
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**

Mazzoni CJ, Ciofi C, Waterhouse RM: **Biodiversity: an atlas of European reference genomes.** *Nature.* 2023; **619**(7969): 252.
PubMed Abstract | **Publisher Full Text**

Nawrocki EP, Eddy SR: **Infernal 1.1: 100-fold faster RNA homology searches.** *Bioinformatics.* 2013; **29**(22): 2933–2935.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Nevers Y, Warwick Vesztrocy A, Rossier V: **Quality assessment of gene repertoire annotations with OMArk.** *Nat Biotechnol.* 2025; **43**(1): 124–133.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Palomba M, Mattiucci S, Crocetta F, *et al.*: **Insights into the role of deep-sea squids of the genus *Histioteuthis* (Histioteuthidae) in the life cycle of ascaridoid parasites in the Central Mediterranean Sea waters.** *Sci Rep.* 2021; **11**(1): 7135.
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**

Thomas KN, Robison BH, Johnsen S: **Two eyes for two purposes: *in situ* evidence for asymmetric vision in the cockeyed squids *Histioteuthis heteropsis* and *Stigmatoteuthis dofleini*.** *Philos Trans R Soc Lond B Biol Sci.* 2017; **372**(1717): 20160069.
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**

Young RE, Vecchione M: ***Stigmatoteuthis arcturi* Robson, 1948.** Version 27 February 2016 (under construction) in The Tree of Life Web Project, http://tolweb.org/, 2016.
Reference Source

Open Peer Review

Current Peer Review Status: ✔ ✔ ?

Version 1

Reviewer Report 08 October 2025

https://doi.org/10.21956/openreseurope.22118.r60082

© **2025 Gerdol M.** 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.

?

Marco Gerdol
University of Trieste, Trieste, Italy

The manuscript by Fernandez-Alvarez and colleagues reports the genome assembly of *Stigmatoteuthis arcturi*, a deep sea squid. Although the approach is technically sound and gold-standard methodologies were used to generate raw sequencing data and to obtain the genome assembly and annotation, a number of points could be slightly improved for the sake of clarity.

Materials and methods: the introductory section may be removed. In understand that this might be a standard sentence recommended for the all ERGA genomes, but in this case non PacBio sequencing was used, for example, and therefore this information, placed here, is somewhat misleading.

"As the squid is not considered an experimental animal" is unclear. Please explain what does this mean. I guess that a blue whale would not be considered an "experimental animal" either, but I strongly doubt that no approval by an ethical committee would be needed in that case. Generally speaking, such approval are linked with the texonomical classification of a species, and regulations about cephalopods may vary from nation to nation. Please be more specific about this point, since the meaning of "experimental animal" is unclear.

Genetic information: as far as I can see from the linked ERGA document, the expected haploid number of chromosomes was 6 (unless there is a mistake in the file), so 46 was the observed number of chromosomes after the assembly. If this was the case, then I don't see any reason for this section to exist or to be placed here: simply report these data later, when the genome assembly statistics are mentioned.

Genome annotation methods: "a select set of clade-specific proteins from UniProt". Please aplain which criteria were used to select these proteins, with particular reference to the fact they were "clade-specific". This does not mean much, unless you explain which taxonomic rank was used to define the clade (e.g. Cephalopoda, Mollusca, etc.).

Genome assembly: I would have expected here a clear statement about the fraction of the

genome assembly size falling within the 46 expected chromosomes.

The organization of Table 2 may look unclear to most readers. The caption should clarify that the consistency scores only apply to OMArk, and that they are linked with all predicted protein-coding gene models (unlink completeness evaluation, which only concerns a subset of genes).

Data availability: a link the gene annotation track should be more explicitly provided in this section, since this is not available from the ERGA portal.

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?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molluscan 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, however I have significant reservations, as outlined above.

Reviewer Report 27 September 2025

https://doi.org/10.21956/openreseurope.22118.r60078

© 2025 Martín-Durán J. 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.

José M Martín-Durán
Queen Mary University of London, London, England, UK

This genome note reports the genome assembly and annotation of the jewelled squid *Stigmatoteuthis arcturi*. This is an important dataset and resource for the study of oceanic cephalopods that will benefit multiple research fields, from phylogenomics and evolutionary biology to ecology. Strengths: This is a gold-standard, high-quality resource. The genome assembly comes with an annotation, which will allow immediate re-use of the data; Weaknesses: none.

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: Evo-Devo, comparative genomics and epigenomics

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 16 September 2025

https://doi.org/10.21956/openreseurope.22118.r60081

© 2025 Yoshida M. 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.

Masa-Aki Yoshida
Shimane University, Shiman, Japan

Overall, the study is well documented and appears publishable; however, for the benefit of future related research, the following minor point should be addressed.
As the authors themselves note, genomic analysis of pelagic squid remains limited and incomplete. Simply estimating genome size by comparison with other species is not acceptable. Although the assembled size appears consistent, both genome size and chromosome number should be independently validated using alternative methods.

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: Bioinformatics, 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.