DATA NOTE

The genome sequence of the Celtic sea slug, *Onchidella celtica* (Audouin & Milne-Edwards, 1832) (Systellommatophora: Onchidiidae)

[version 1; peer review: 2 approved]

Nova Mieszkowska[1], Marine Biological Association Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics team, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

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

v1 **First published:** 06 Oct 2025, **10**:557
https://doi.org/10.12688/wellcomeopenres.25000.1
Latest published: 06 Oct 2025, **10**:557
https://doi.org/10.12688/wellcomeopenres.25000.1

Abstract

We present a genome assembly from an individual *Onchidella celtica* (Celtic sea slug; Mollusca; Gastropoda; Systellommatophora; Onchidiidae). The genome sequence has a total length of 1 141.02 megabases. Most of the assembly (99.92%) is scaffolded into 18 chromosomal pseudomolecules. Gene annotation of this assembly on Ensembl identified 12 582 protein-coding genes. The mitochondrial genome has also been assembled, with a length of 13.99 kilobases. This assembly was generated as part of the Darwin Tree of Life project, which produces reference genomes for eukaryotic species found in Britain and Ireland.

Keywords

Onchidella celtica; Celtic sea slug; genome sequence; chromosomal; Systellommatophora

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 06 Oct 2025	✓ view	✓ view

1. **Dr. Sujan Kumar Datta**, University of Dhaka, Dhaka, Bangladesh

2. **Erich M. Schwarz**, Cornell University, Ithaca, 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: Mieszkowska N: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (220540) and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Mieszkowska N, Marine Biological Association Genome Acquisition Lab, Darwin Tree of Life Barcoding Collective *et al.* **The genome sequence of the Celtic sea slug, *Onchidella celtica* (Audouin & Milne-Edwards, 1832) (Systellommatophora: Onchidiidae) [version 1; peer review: 2 approved]** Wellcome Open Research 2025, **10**:557
https://doi.org/10.12688/wellcomeopenres.25000.1

First published: 06 Oct 2025, **10**:557 https://doi.org/10.12688/wellcomeopenres.25000.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Spiralia; Lophotrochozoa; Mollusca; Gastropoda; Heterobranchia; Euthyneura; Panpulmonata; Eupulmonata; Systellommatophora; Onchidioidea; Onchidiidae; *Onchidella*; *Onchidella celtica* (Audouin & Milne-Edwards, 1832) (NCBI: txid36933)

Background

Onchidella celtica (Audouin & Milne-Edwards, 1832) is a small (up to 12 mm in length and 6 mm width) marine pulmonate gastropod mollusc from the family Onchidiidae. Its recorded distribution range extends from Mauritania to northwest Scotland, being predominantly found on the western coastlines of the UK (*Onchidella celtica* distribution map. World Register of Marine Species, 2025; *Onchidella celtica*. NBN Atlas species page, 2025). It occurs on exposed rocky shores (Forbes *et al.*, 1853; Jeffreys, 1867; Joyeux-Laffuie, 1882), living gregariously amongst mussels, barnacles, and in rock crevices (Beauchamp, 1923; Fretter, 1943; Russell, 1925). It forages when the tide is out, feeding on juvenile algae and diatoms (Fretter, 1943; Kent & Hawkins, 2019), and is more commonly observed in damp, humid conditions (Mieszkowska pers. obs), likely accounting for the observed locally patchy distribution (Dron, 1971; Joyeux-Laffuie, 1882; Russell, 1925). Various reports of the vertical distribution of *O. celtica* exist within the literature (Jeffreys, 1867), however, more recent work favours a vertical range coinciding with that of neap tides and possibly a little higher vertically than the high neap tide level (Britton, 1982; Dron, 1971).

This air-breathing sea slug has an oval body with a mantle that is black to dark green in colour and is covered in large tubercles. There are two short, thick, cylindrical tentacles with eyes at the tips that are located on the head (Rowley, 2005). As a pulmonate gastropod, it has no gills, with the mantle cavity acting as a lung (Barfield, 2003). *O. celtica* is a hermaphrodite, laying from 60 to 100 eggs in tubular capsules approximately 1 cm in diameter, and undergoing reciprocal fertilisation (Barfield, 2003).

We present a chromosome-level genome sequence for *Onchidella celtica*, the first high-quality genome for the genus *Onchidella* (data obtained via NCBI datasets, O'Leary *et al.*, 2024). The assembly was produced using the Tree of Life pipeline from a specimen collected in Trevone, Cornwall, United Kingdom (Figure 1). This assembly was generated as part of the Darwin Tree of Life Project, which aims to generate high-quality reference genomes for all named eukaryotic species in Britain and Ireland to support research, conservation, and the sustainable use of biodiversity (Blaxter *et al.*, 2022).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult *Onchidella celtica* (specimen ID MBA-210329-004A, ToLID xgOncCelt3; Figure 1), collected from Trevone, Cornwall, United Kingdom (latitude 50.545, longitude –4.985) on 2021-03-29.

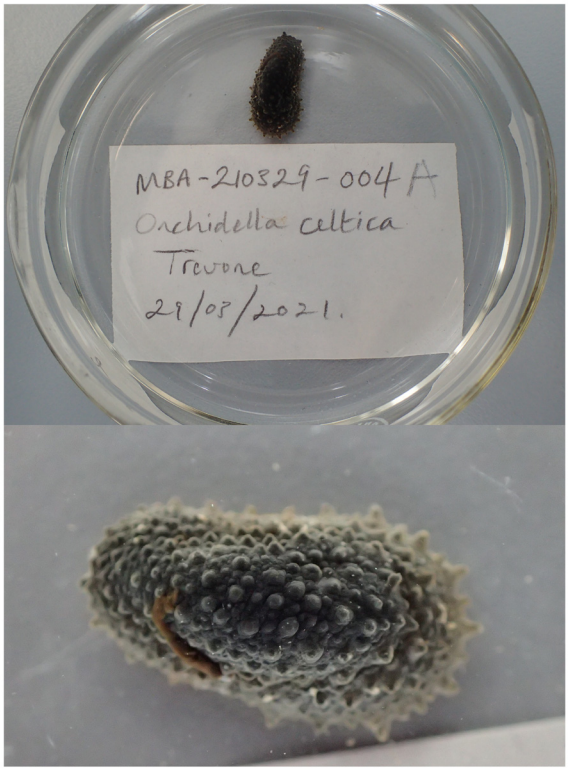

Figure 1. Photographs of the *Onchidella celtica* (xgOncCelt3) specimen used for genome sequencing.

The specimen was collected and identified by Nova Mieszkowska (Marine Biological Association). For the Darwin Tree of Life sampling and metadata approach, refer to Lawniczak *et al.* (2022).

The initial identification was verified by an additional DNA barcoding process according to the framework developed by Twyford *et al.* (2024). A small sample was dissected from the specimen and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI) (see the protocol). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding are available on protocols.io.

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 xgOncCelt3 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the anterior body was homogenised by powermashing using

a PowerMasher II tissue disruptor. 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 29.2 ng/μL and a yield of 3 796.00 ng.

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 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 on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 μL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25M SMRT cell. The SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen anterior body tissue of the xgOncCelt3 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.

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 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).

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 18 breaks, 35 joins, and removal of 1 haplotypic duplication. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation. PretextSnapshot 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 the 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 containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

PacBio sequencing of the *Onchidella celtica* specimen generated 81.19 Gb (gigabases) from 9.30 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 1 068.34 Mb, with a heterozygosity of 0.47% and repeat content of 31.83% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 71× coverage. Hi-C sequencing produced 438.64 Gb from 2 904.90 million reads, which were used to scaffold the assembly. 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 1 141.02 Mb in 114 scaffolds, with 923 gaps, and a scaffold N50 of 61.4 Mb (Table 2).

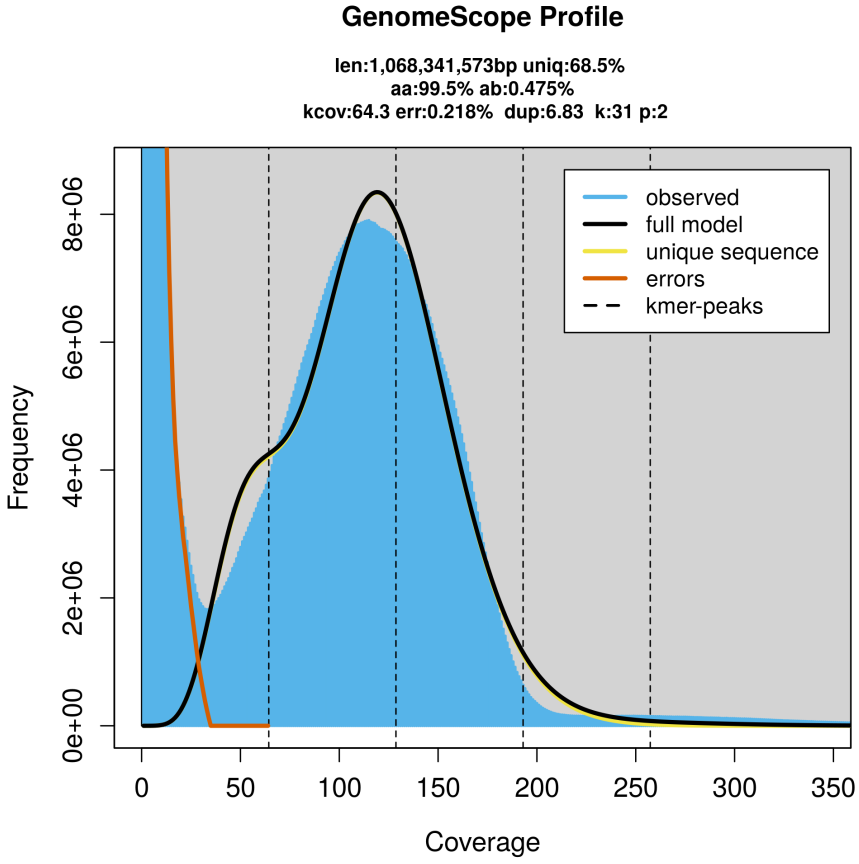

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Most of the assembly sequence (99.92%) was assigned to 18 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.

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 56.0. The *k*-mer completeness is 97.06% for the primary assembly, 33.36% for the alternate haplotype, and 97.98% for the combined assemblies (Figure 4).

BUSCO v.5.5.0 analysis using the mollusca_odb10 reference set (n = 5 295) identified 93.6% of the expected gene set (single = 92.8%, duplicated = 0.8%). 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.

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 **6.C.Q56**, meeting the recommended reference standard.

Genome annotation report

The *Onchidella celtica* genome assembly (GCA_963931925.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI), using the Ensembl genebuild for non-vertebrate species pipeline. This annotation includes 22 534 transcribed

Table 1. Specimen and sequencing data for BioProject PRJEB71538.

Platform	PacBio HiFi	Hi-C
ToLID	xgOncCelt3	xgOncCelt3
Specimen ID	MBA-210329-004A	MBA-210329-004A
BioSample (source individual)	SAMEA110449713	SAMEA110449713
BioSample (tissue)	SAMEA110450538	SAMEA110450538
Tissue	anterior body	anterior body
Instrument	Revio	Illumina NovaSeq 6000
Run accessions	ERR12408789; ERR12408788	ERR12512733
Read count total	9.30 million	2 904.90 million
Base count total	81.19 Gb	438.64 Gb

Table 2. Genome assembly statistics.

Assembly name	xgOncCelt3.1
Assembly accession	GCA_963931925.1
Alternate haplotype accession	GCA_963931835.1
Assembly level	chromosome
Span (Mb)	1 141.02
Number of chromosomes	18
Number of contigs	1 037
Contig N50	1.96 Mb
Number of scaffolds	114
Scaffold N50	61.4 Mb
Organelles	Mitochondrion: 13.99 kb

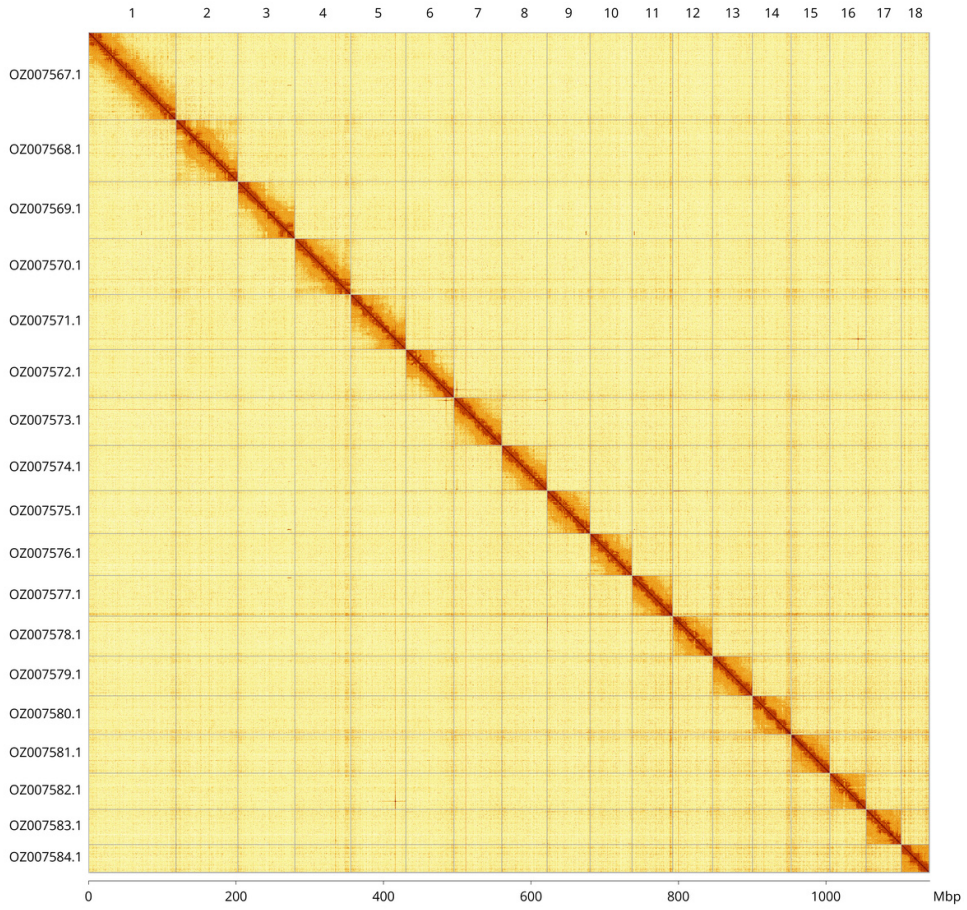

Figure 3. Hi-C contact map of the *Onchidella celtica* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Onchidella celtica* xgOncCelt3.

INSDC accession	Molecule	Length (Mb)	GC%
OZ007567.1	1	118.62	39.50
OZ007568.1	2	83.86	40
OZ007569.1	3	77.53	40
OZ007570.1	4	75.73	40
OZ007571.1	5	74.65	40
OZ007572.1	6	65.26	40
OZ007573.1	7	64.92	40.50
OZ007574.1	8	61.40	40
OZ007575.1	9	58.44	40
OZ007576.1	10	56.89	40.50
OZ007577.1	11	55.03	40.50
OZ007578.1	12	54.29	40.50
OZ007579.1	13	53.82	40.50
OZ007580.1	14	52.58	40.50
OZ007581.1	15	52.35	40.50
OZ007582.1	16	49.40	40.50
OZ007583.1	17	47.46	40.50
OZ007584.1	18	37.82	40.50

mRNAs from 12 582 protein-coding and 3 508 non-coding genes. The average transcript length is 14 864.72 bp, with an average of 1.40 coding transcripts per gene and 7.00 exons per transcript. For further information about the annotation, please refer to the annotation page on Ensembl.

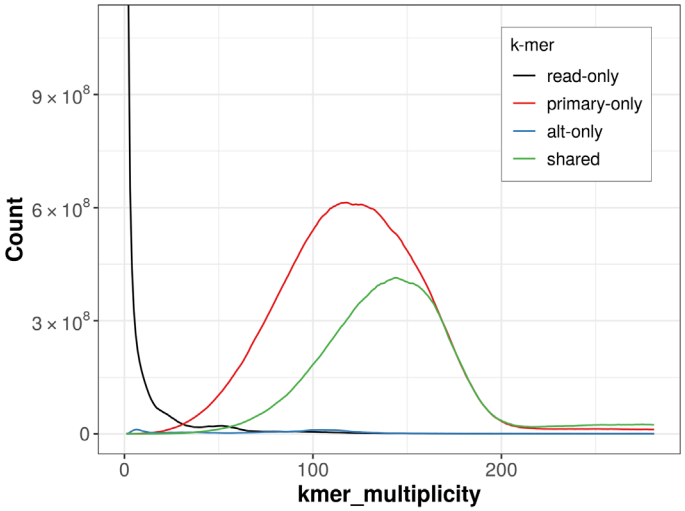

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers 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.

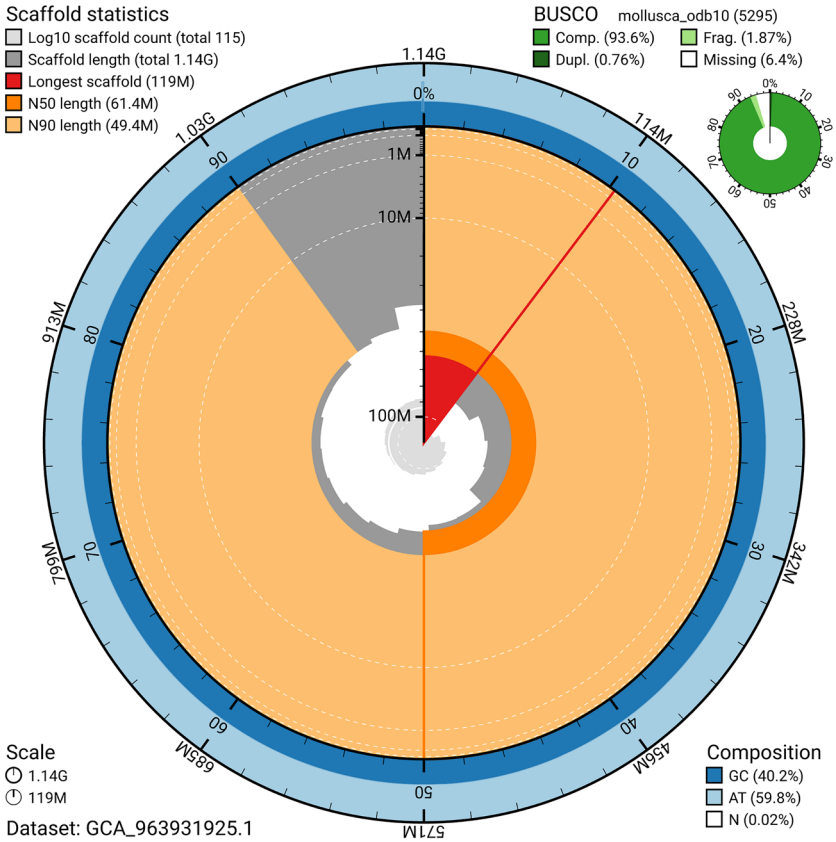

Figure 5. Assembly metrics for xgOncCelt3.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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 mollusca_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the BlobToolKit viewer.

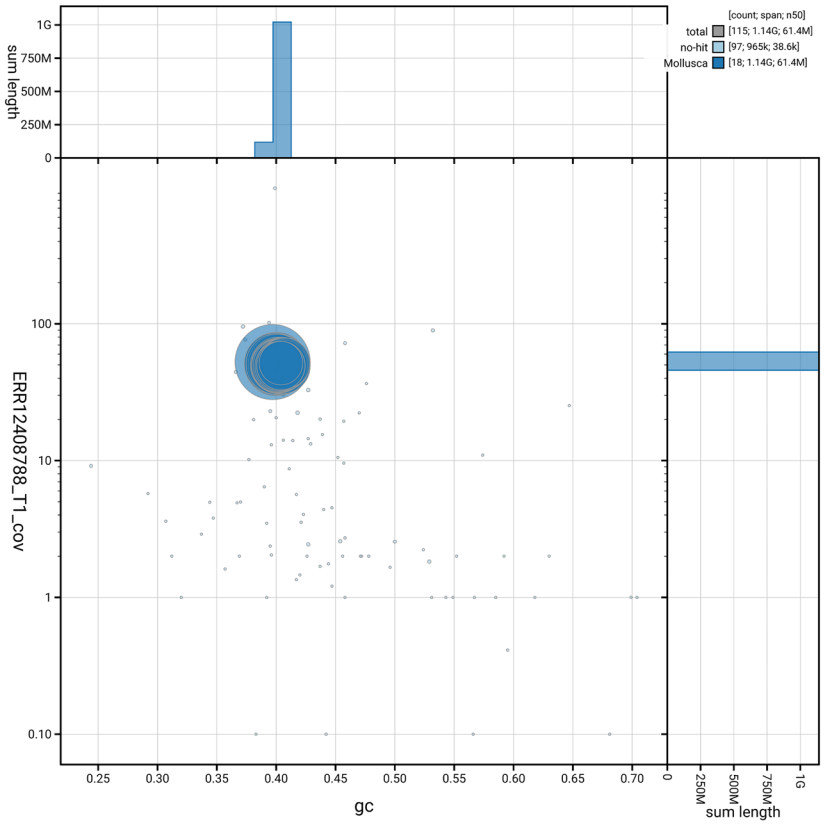

Figure 6. BlobToolKit GC-coverage plot for xgOncCelt3.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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 *Onchidella celtica* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q56	6.C.Q40
Contig N50 length	1.96 Mb	≥ 1 Mb
Scaffold N50 length	61.40 Mb	= chromosome N50
Consensus quality (QV)	Primary: 56.8; alternate: 54.9; combined: 56.0	≥ 40
k-mer completeness	Primary: 97.06%; alternate: 33.36%; combined: 97.98%	≥ 95%
BUSCO	C:93.6% [S:92.8%; D:0.8%]; F:1.9%; M:4.5%; n:5 295	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.92%	≥ 90%

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website. 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: Onchidella celtica (Celtic sea slug). Accession number PRJEB71538. The genome sequence is released openly for reuse. The *Onchidella celtica* genome sequencing initiative is part of the Darwin Tree of Life Project (PRJEB40665) and the Sanger Institute Tree of Life Programme (PRJEB43745). 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

Contributors are listed at the following links:

- Members of the Marine Biological Association Genome Acquisition Lab
- Members of the Darwin Tree of Life Barcoding collective
- 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/tolkit/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
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass

Software	Version	Source
MerquryFK	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	23.04.1	https://github.com/nextflow-io/nextflow
PretextSnapshot	N/A	https://github.com/sanger-tol/PretextSnapshot
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
purge_dups	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	0.4.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.4.0	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

Barfield P: **Notes on the natural history of the Celtic sea-slug, *Onchidella celtica* (Cutier, 1817).** *Porcupine Marine Natural History Society Newsletter.* 2003; (13): 12–14.
Reference Source

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

Beauchamp PM: **Études de bionomie intercotidale: les îles de Ré et d'Yeu.** *Archives de Zoologie Expérimentale et Générale.* 1923; **61**: 455–520.

Blaxter M, Mieszkowska N, Di Palma F, *et al.*: **Sequence locally, think globally: the Darwin Tree of Life project.** *Proc Natl Acad Sci U S A.* 2022; **119**(4): e2115642118.
PubMed Abstract | Publisher Full Text | Free Full Text

Britton KM: **Studies on the order Onchidiacea (Mollusca, Gastropoda): With special reference to the species from Great Britain and Hong Kong.** 1982; 123.

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

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

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

Dron EM: **Aspects of the physiological ecology and behaviour of *Onchidella celtica* (Forbes and Hanley, 1852), with a comparative account of the light reactions of *Onchidella floridana* (Dall.).** 1971; 82.

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

Forbes E, Hanley SCT, Hanley S: **A history of British Mollusca, and their shells.** John Van Voorst, 1853; **1**.
Reference Source

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

Fretter V: **Studies in the functional morphology and embryology of *Onchidella celtica* (Forbes and Hanley) and their bearing on its relationships.** *J Mar Biol Assoc U K.* 1943; **25**(4): 685–720.
Publisher Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | Publisher Full Text | Free Full Text

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): giaa153.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Jeffreys JG: **British conchology: or an account of the Mollusca which now inhabit the British Isles and the surrounding seas.** J. van Voorst, 1867; **4**.
Publisher Full Text

Joyeux-Laffuie J: **Organisation et développement de l'Oncidie, Oncidium celticum Cuv.** *Archives de Zoologie Expérimentale et Générale.* 1882.
Reference Source

Kent RML, Hawkins SJ: **On the timing and duration of foraging in *Onchidella celtica*.** *J Mar Biol Assoc U K.* 2019; **99**(2): 411–419.
Publisher Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

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**

Lawniczak 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**

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.
Reference Source

O'Leary NA, Cox E, Holmes JB, *et al.*: **Exploring and retrieving sequence and metadata for species across the Tree of Life with NCBI datasets.** *Sci Data.* 2024; **11**(1): 732.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

***Onchidella celtica* distribution map.** World Register of Marine Species, 2025.
Reference Source

***Onchidella celtica*.** NBN Atlas species page, 2025.
Reference Source

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**

Rowley SJ: ***Onchidella celtica* Celtic sea slug.** 2005.
Reference Source

Russell FS: **On the occurrence of *Onchidella celtica (Cuvier)* on the Cornish coast.** *J Mar Biol Assoc U K.* 1925; **13**(4): 981–982.
Publisher 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**

Twyford AD, Beasley J, Barnes I, *et al.*: **A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life project [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2024; **9**: 339.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

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 27 December 2025

https://doi.org/10.21956/wellcomeopenres.27538.r140727

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

This data note describes the first chromosomal-quality genome assembly for Onchidella celtica, with a useful set of gene predictions by Ensembl. The general process of genome assembly, curation, and annotation is precisely described using methods that are state-of-the-art. This text has had a previous review by Dr. Sujan Kumar Datta that raised reasonable points, and it would appear that this version is intended to address those points.

1. It would appear that Figure 5 is meant to address Dr. Datta's question about contamination or non-target sequences. However, I cannot see any discussion of the biological interpretation or practical response to Figure 5's data in the text; there is a section of Methods describing how Figure 5 was generated, but nothing I can find about what it means or what was done about it. Could the authors please add these?

2. No comparison of genome assembly traits to other gastropods or pulmonates was provided. However, this may not be necessary since the standard format of genome reports by WTSI's Tree of Life generally does not seek to provide such comparisons. Therefore, I believe this point can be ignored.

3. Figures 2 and 3 now appear to me to be fully acceptable.

4. The authors have added a pleasingly informative paragraph about the known biology of Onchidella celtica. More generally, the text reads well, with no obvious errors.

In addition to the points raised by Dr. Datta, I see two other minor points that the authors may wish to address.

5. In the Methods, the authors refer first to "a PowerMasher II tissue disruptor" (for nucleic acid extraction) and then to "the Diagnocine Power Masher-II" (for Hi-C). Unless these really were two different pieces of laboratory equipment, could the authors please pick a single set of words to describe this piece of machinery and use it in both places?

6a. In the title and elsewhere in the text, the authors describe this species as "Onchidella celtica (Audouin & Milne-Edwards, 1832)". However, I noticed that among the references they cite, this species went several different taxonomic designations: "Onchidella celtica (Cutler, 1817)" [Barfield, 2003]; "Onchidella celtica (Forbes and Hanley, 1852)" [Dron, 1971]; and "Onchidella celtica (Cuvier)" [Russell, 1925]. Which of these is correct? I believe the "(Audouin & Milne-Edwards, 1832)" designation is actually correct; could the authors cite a reference for it? Perhaps this would be a valid reference: Audouin, V.; Milne Edwards, H. (1832-1834). Recherches pour servir à l'histoire naturelle du littoral de la France. Paris: Crochard, 2 vols [vol. 1, 406 pp.; vol. 2, 290 pp., 8 pls].

6b. Also, should this be written as "Audouin & Milne-Edwards, 1832" or "Audouin and Milne-Edwards, 1832"?

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 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 06 November 2025

https://doi.org/10.21956/wellcomeopenres.27538.r135438

Dr. Sujan Kumar Datta
Department of Zoology, University of Dhaka, Dhaka, Dhaka Division, Bangladesh

Summary
This Data Note reports a high-quality, chromosome-level genome assembly and annotation of Onchidella celtica, the Celtic sea slug. Conducted within the framework of the Darwin Tree of Life (DToL) project, this work provides a valuable genomic resource for studies in molluscan

systematics and evolution. It represents the first reference genome for the genus Onchidella, addressing a key gap in molluscan genomic resources. The assembly demonstrates excellent data quality (BUSCO 93.6%, QV 56, and 99.92% chromosomal assignment) and includes comprehensive documentation of protocols, software versions, and accession numbers. The manuscript is clearly written, methodologically rigorous, and adheres to the principles of open data and reproducibility.

Areas for Improvement
1. The text could briefly mention whether any contamination or non-target sequences were identified and removed during curation using the ASCC pipeline, as this would strengthen confidence in the overall data quality.
2. It would also be useful to include a short comparison of the genome size, GC content, or gene count relative to other gastropods or pulmonate species (e.g., *Onchidella floridana* or *Aplysia* species) to provide additional context and highlight the distinct features of this assembly.
3. Figures: Figure 2 (k-mer plot)- label homozygous and heterozygous peaks for clarity. Include a short explanatory sentence in the legend of Figure 3 describing the main chromosomal pattern visible in the Hi-C map.
4. Minor Textual Improvements: -Ensure consistent italicization of all species names.
-Replace “megabases” with “Mb” throughout for consistency with other Tree of Life notes.
-Consider a one-sentence summary of the ecological or conservation relevance of *O. celtica* (air-breathing, intertidal adaptation, patchy distribution).
-In the second paragraph of the Genome Assembly section, an extra hyphen (“-”) has been inserted. Please remove it.
- Remove the last citation (Blaxter et al., 2022) from the background section.
- The article “The” can be omitted from the title as “Genome sequence of the Celtic sea slug, *Onchidella celtica* (Audouin & Milne-Edwards, 1832) (Systellommatophora: Onchidiidae).”
Recommendation
This is a technically sound and well-documented genome report. With minor clarifications to figures and a short comparative context, it will make a valuable addition to the *Darwin Tree of Life* genome collection.

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: Fisheries (Marine and Freshwater fishes, Crustacea, Mollusk), Aquaculture, Marine Biotechnology, Marine Products, Molecular Biology, Stock assessment

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.