



DATA NOTE

The genome sequence of the China Limpet, *Patella*

ulyssiponensis Gmelin, 1791 (Patellidae)

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

Nova Mieszkowska ^{1,2}, Stephen J Hawkins^{1,3}, Helena B S M Côrte-Real², Louise B Firth^{4,5}, 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

¹The Marine Biological Association, Plymouth, England, UK²University of Liverpool, Liverpool, England, UK³University of Southampton, Southampton, England, UK⁴University of Cork, Cork, Ireland⁵University of Plymouth, Plymouth, England, UK

V1 First published: 11 Aug 2025, 10:419
<https://doi.org/10.12688/wellcomeopenres.24681.1>

Latest published: 11 Aug 2025, 10:419
<https://doi.org/10.12688/wellcomeopenres.24681.1>

Abstract

We present a genome assembly from a specimen of *Patella ulyssiponensis* (China Limpet; Mollusca; Gastropoda; Patellidae). The genome sequence has a total length of 693.56 megabases. Most of the assembly (99.78%) is scaffolded into 8 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 14.94 kilobases. Gene annotation of this assembly on Ensembl identified 21 151 protein-coding genes.

Keywords

Patella ulyssiponensis, China Limpet, genome sequence, chromosomal, Mollusca



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

Open Peer Review

Approval Status

	1	2	3	4
version 1				
11 Aug 2025	view	view	view	view

1. **Ximing Guo** , Rutgers University New Brunswick, New Brunswick, USA
2. **Masa-Aki Yoshida** , Shimane University, Matsue, Shiman,, Japan
3. **José M Martín-Durán** , Queen Mary University of London, London, UK
4. **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.

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; **Hawkins SJ:** Writing – Original Draft Preparation; **Côrte-Real HBSM:** Writing – Original Draft Preparation; **Firth LB:** Writing – Original Draft Preparation;

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.

Copyright: © 2025 Mieszkowska N *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: Mieszkowska N, Hawkins SJ, Côrte-Real HBSM *et al.* **The genome sequence of the China Limpet, *Patella ulyssiponensis* Gmelin, 1791 (Patellidae) [version 1; peer review: 2 approved, 2 approved with reservations]** Wellcome Open Research 2025, 10:419 <https://doi.org/10.12688/wellcomeopenres.24681.1>

First published: 11 Aug 2025, 10:419 <https://doi.org/10.12688/wellcomeopenres.24681.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Spiralia; Lophotrochozoa; Mollusca; Gastropoda; Patellogastropoda; Patelloidea; Patellidae; *Patella*; *Patella ulyssiponensis* Gmelin, 1791 (NCBI:txid334708)

Background

Patella ulyssiponensis Gmelin, 1791 (often referred to as *P. aspera* Röding, 1798) occurs on exposed rocky shores from Mauritania to Norway, including all suitable shores of Ireland and the British Isles (except the soft coasts between Kent and Flamborough Head) and in the Black Sea and Mediterranean. Its wide range led to extensive synonymy and shifting names based on shell characters (Christiaens, 1973 listed 53). Early names (*P. ulyssiponensis* Gmelin, 1791; *P. aspera* Röding, 1798; *P. athletica* Bean, 1844) were distinguished or merged by 19th- and early 20th-century authors (Dautzenberg, 1906; Jeffreys, 1865; Powell, 1973). Genetic methods – from allozyme electrophoresis (Gaffney, 1980) to DNA sequencing (Koufopanou *et al.*, 1999; Sá-Pinto *et al.*, 2005; Sá-Pinto *et al.*, 2012; Weber & Hawkins, 2005) – have confirmed that *P. ulyssiponensis* (European mainland) is distinct from *P. aspera* (Azores, Canaries, Madeira). The status of Black Sea and southern Atlantic (St Helena, Angola) populations remains uncertain.

Patella ulyssiponensis occupies low-shore open rock and mid-shore pools, often where crustose calcareous algae (CCA) grow (Delany *et al.*, 1998; Firth & Crowe, 2008). On very exposed shores it reappears under the *Laminaria digitata* canopy, where wave action limits turf algae (Hawkins & Harkin, 1985). Limpet grazing maintains bare patches on CCA, preventing turfs and possibly aiding CCA growth; larvae may preferentially settle on CCA (Bowman, 1981; Firth *et al.*, 2023).

This species is a protandrous hermaphrodite, maturing first as male and often later changing to female (Thompson, 1980). In related species sex change can depend on density; this has yet to be tested in *P. ulyssiponensis*. In northern Portugal there are two spawning peaks (September to January and March to June, Ribeiro *et al.*, 2009), whereas in Ireland and the British Isles there is a single annual peak, with gamete maturation in early summer and spawning in October (Thompson, 1979).

Transplant experiments show that *P. ulyssiponensis* can shift red algal turf limits upward on the shore (Boaventura *et al.*, 2002). In Portugal it co-exists in pools with juvenile *P. depressa* and the pulmonate *Siphonaria pectinata*, which later move onto open rock; by contrast, *P. ulyssiponensis* remains in pools, controlling macroalgae and keeping CCA free of epibionts (Firth *et al.*, 2023; O'Connor & Crowe, 2005).

Beyond grazing, this limpet bioerodes softer rock, creating microhabitats that support diverse species (Cooper *et al.*, 2025). Its empty shells also shelter macroalgae from other grazers. Together, these activities make *P. ulyssiponensis* an ecosystem engineer with both habitat-forming and habitat-modifying roles.

The genome of *Patella ulyssiponensis* presented here was assembled using the Tree of Life pipeline from a specimen collected in Church Reef, Wembury, Devon, United Kingdom (Figure 1).

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult *Patella ulyssiponensis* (specimen ID MBA-210201-002A, ToLID xgPatUlys2; Figure 1), collected from Church Reef, Wembury, Devon, United Kingdom (latitude 50.3155, longitude -4.0829) on 2021-02-01. The specimen was collected and identified by Nova Mieszkowska (Marine Biological Association). A second specimen was used for Hi-C and RNA sequencing (specimen ID MBA-200706-001A, ToLID xgPatUlys1). It was collected from Godrevy, Cornwall, United Kingdom (latitude 50.2379, longitude -5.3961) on 2020-07-06. The specimen was collected and identified by Nova Mieszkowska and Rob Mrowicki.

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). 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

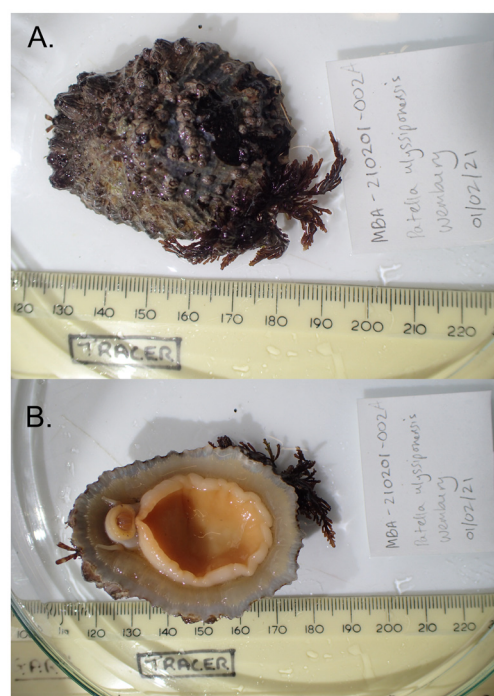


Figure 1. Photograph of the *Patella ulyssiponensis* (xgPatUlys2) specimen from which samples were taken for genome sequencing.

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 xgPatUlys2 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue from the muscle was homogenised by [cryogenic disruption](#) using the Covaris cryoPREP® Automated Dry Pulverizer. HMW DNA was extracted using the [Automated MagAttract v2](#) protocol. DNA was sheared into an average fragment size of 12–20 kb following the [Megaruptor®3 for LI PacBio](#) protocol. Sheared DNA was purified by [automated SPRI](#) (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from muscle tissue of xgPatUlys1 in the Tree of Life Laboratory at the WSI using the [RNA Extraction: Automated MagMax™ mirVana protocol](#). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

Library preparation and sequencing were performed at the WSI Scientific Operations core. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California, USA), following the manufacturer's instructions. The kit includes the reagents required for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead cleanup, and nuclease treatment. Size-selection and clean-up were carried out using diluted AMPure PB beads (Pacific Biosciences). DNA concentration was quantified using the Qubit Fluorometer v4.0 (ThermoFisher Scientific) with Qubit 1X dsDNA HS assay kit and the final library fragment size analysis was carried out using the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) and the gDNA 55kb BAC analysis kit.

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

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen muscle tissue of the xgPatUlys1 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 Diagenode Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

Prior to sequencing, libraries were normalised to 10 ng/μL. Normalised libraries were quantified again and equimolar and/or weighted 2.8 nM pools. Pool concentrations were checked using the Agilent 4200 TapeStation (Agilent) with High Sensitivity D500 reagents before sequencing. Sequencing was performed using paired-end 150 bp reads on the Illumina NovaSeq 6000.

RNA library preparation and sequencing

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

Genome assembly

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

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were

identified and removed using `purge_dups` (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using `bwa-mem2` (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with the `--break` 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 32 breaks, 29 joins, and removal of 12 haplotypic duplications. The curation process is documented at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextViewSnapshot was used to generate a Hi-C contact map of the final assembly.

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) 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 BlobToolKit pipeline (Challis *et al.*, 2020). The pipeline aligns PacBio reads using `minimap2` (Li, 2018) and `SAMtools` (Danecek *et al.*, 2021) to generate coverage tracks. Simultaneously, it queries the GoAT database (Challis *et al.*, 2023) to identify relevant BUSCO lineages and runs BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using `DIAMOND blastp` (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with `DIAMOND blastx`. Sequences without hits are chunked using `seqtk` and aligned to the NT database with `blastn` (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a `blobdir` for visualisation. The BlobToolKit pipeline was developed using `nf-core` tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with package management via Conda and Bioconda (Grüning *et al.*, 2018), and containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Genome sequence report

Sequence data

The genome of a specimen of *Patella ulyssiponensis* was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 23.58 Gb (gigabases) from 2.68 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 685.09 Mb, with a heterozygosity of 1.45% and repeat content of 36.76% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 33× coverage. Hi-C sequencing produced 96.42 Gb from 638.57 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The final assembly has a total length of 693.56 Mb in 27 scaffolds, with 73 gaps, and a scaffold N50 of 86.69 Mb (Table 2).

Most of the assembly sequence (99.78%) was assigned to 8 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 64.2. The *k*-mer completeness is 72.96% for the primary assembly, 72.67% for the alternate haplotype, and 98.80% for the combined assemblies (Figure 4). BUSCO analysis using the mollusca_odb10 reference set (*n* = 5 295) identified 88.5% of the expected gene set (single = 87.5%, duplicated = 1.0%). 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 **7.C.Q63**, meeting the recommended reference standard.

Genome annotation report

The *Patella ulyssiponensis* genome assembly (GCA_963678685.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 61 607 transcribed mRNAs from 21 151 protein-coding and 22 298 non-coding genes. The average transcript length is 14 847.99 bp. There are

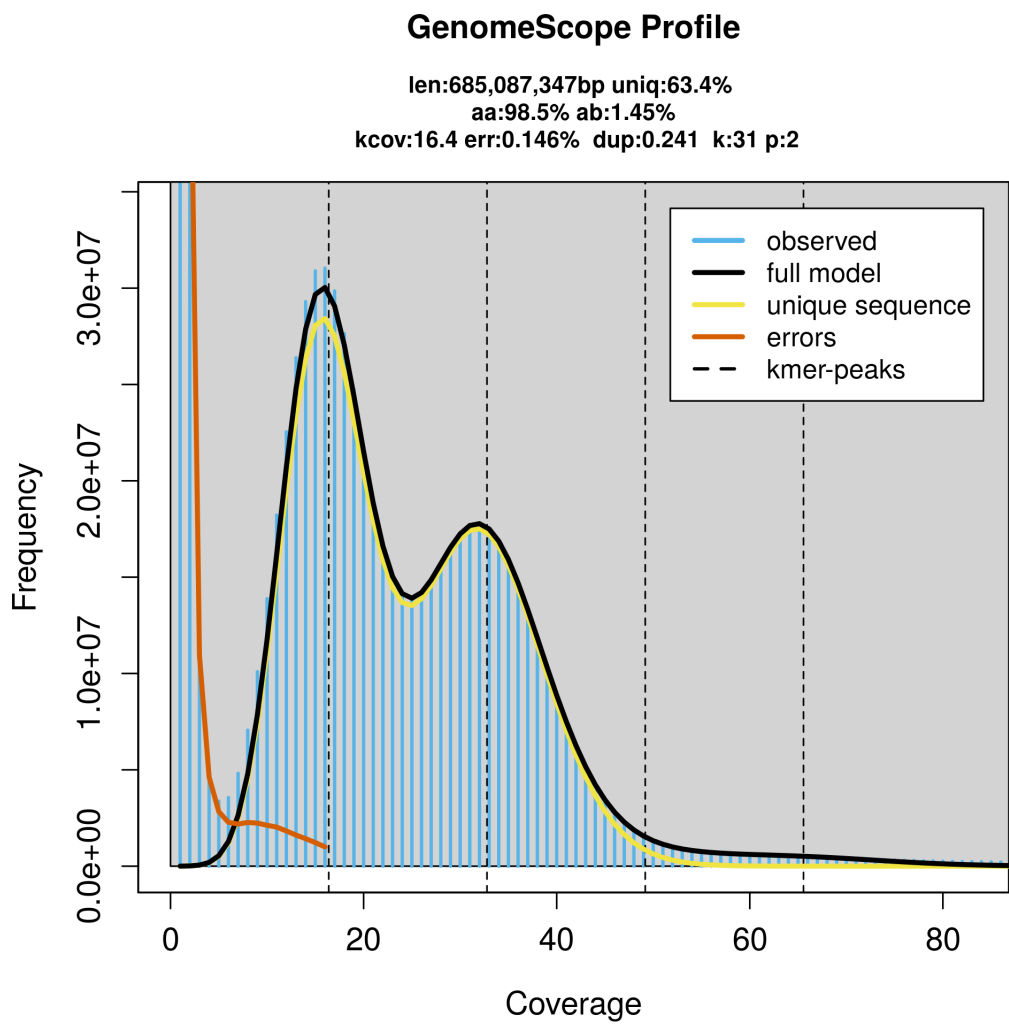


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB63446.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	xgPatUlys2	xgPatUlys1	xgPatUlys1
Specimen ID	MBA-210201-002A	MBA-200706-001A	MBA-200706-001A
BioSample (source individual)	SAMEA8419699	SAMEA7536242	SAMEA7536242
BioSample (tissue)	SAMEA8419982	SAMEA7536328	SAMEA7536329
Tissue	muscle	muscle	muscle
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR11593810; ERR11593809	ERR11606324	ERR11606323
Read count total	2.68 million	638.57 million	44.63 million
Base count total	23.58 Gb	96.42 Gb	6.74 Gb

Table 2. Genome assembly statistics.

Genome assembly	Primary assembly
Assembly name	xgPatUlys2.1
Assembly accession	GCA_963678685.1
Alternate haplotype accession	GCA_963678665.1
Assembly level	chromosome
Span (Mb)	693.56
Number of chromosomes	8
Number of contigs	100
Contig N50	16.75 Mb
Number of scaffolds	27
Scaffold N50	86.69 Mb
Longest scaffold length (Mb)	104.01
Organelles	Mitochondrial genome: 14.94 kb

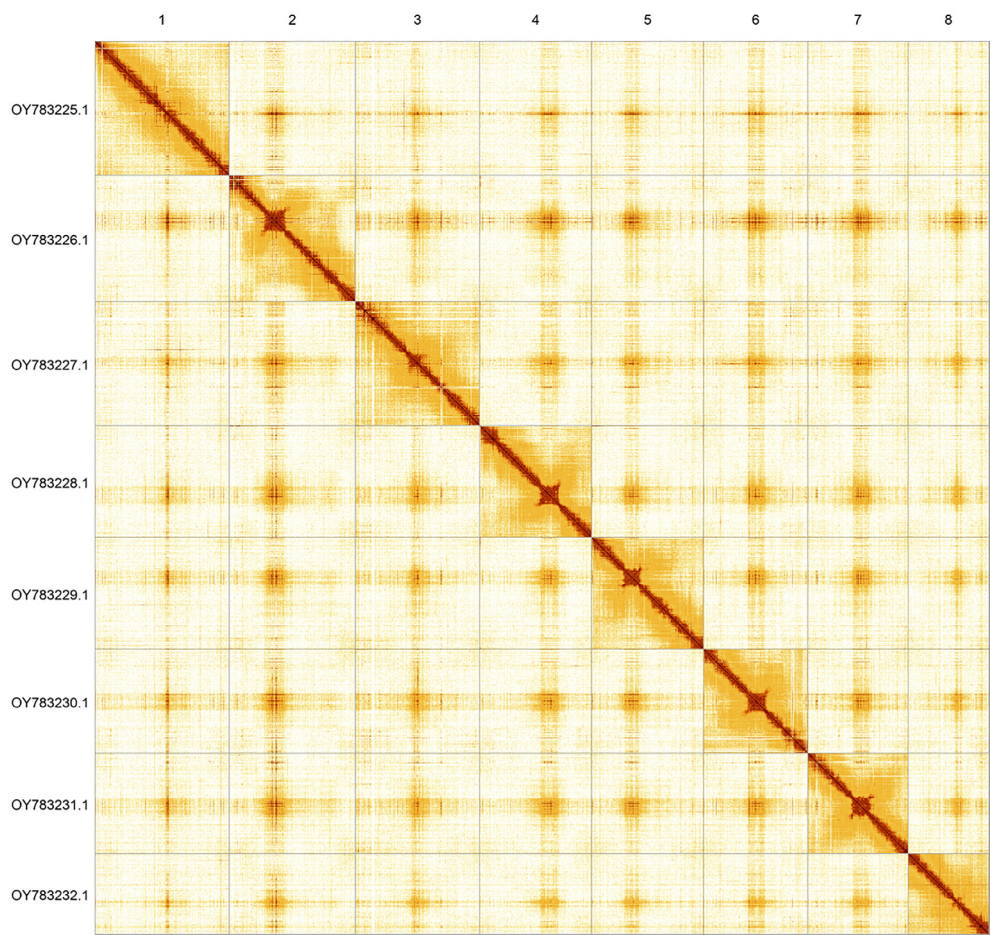


Figure 3. Hi-C contact map of the *Patella ulyssiponensis* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Patella ulyssiponensis* xgPatUlys2.

INSDC accession	Molecule	Length (Mb)	GC%
OY783225.1	1	104.01	36.50
OY783226.1	2	97.66	36
OY783227.1	3	96.22	36.50
OY783228.1	4	86.69	36
OY783229.1	5	86.65	36
OY783230.1	6	80.63	36.50
OY783231.1	7	77.60	36.50
OY783232.1	8	62.57	36.50

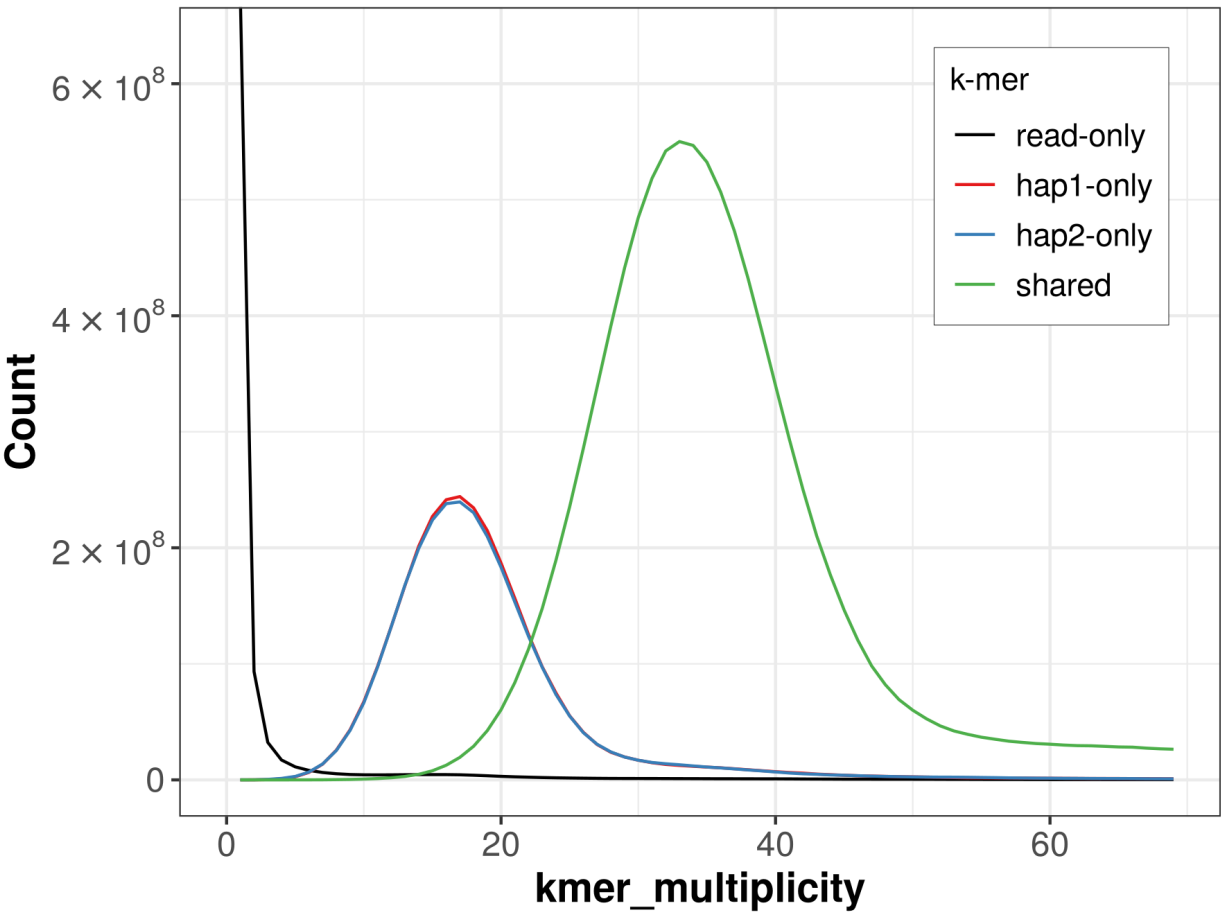


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

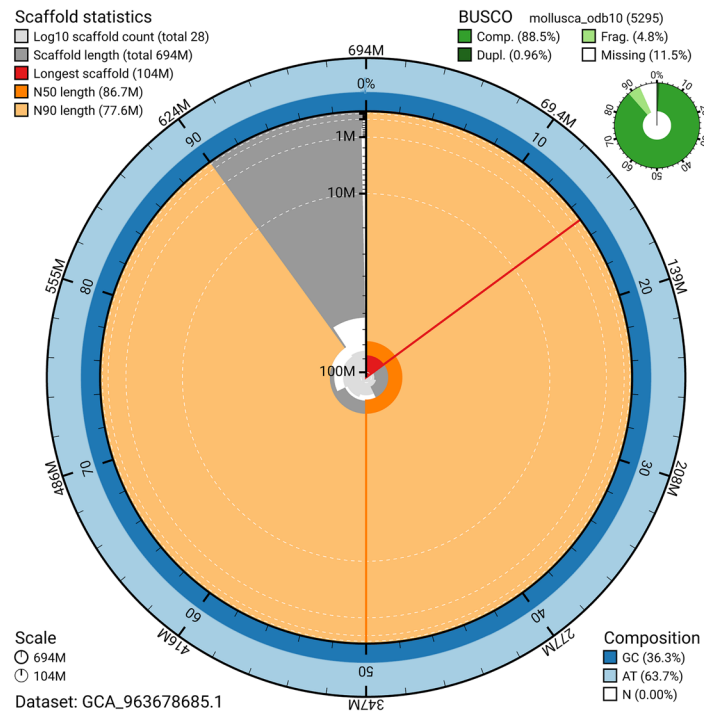


Figure 5. Assembly metrics for xgPatUlys2.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. The outermost blue 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](#).

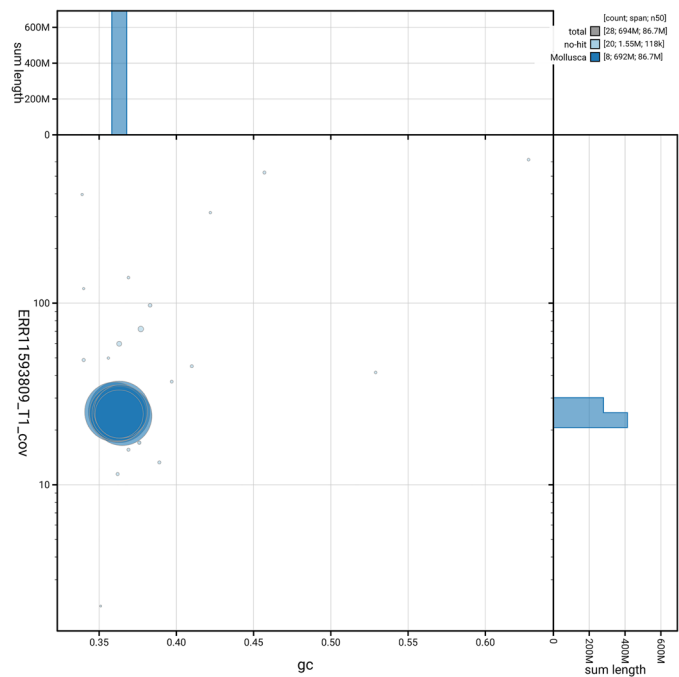


Figure 6. BlobToolKit GC-coverage plot for xgPatUlys2.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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 *Patella ulyssiponensis* assembly.

Measure	Value	Benchmark
EBP summary	7.C.Q63	6.C.Q40
Contig N50 length	16.75 Mb	≥ 1 Mb
Scaffold N50 length	86.69 Mb	= chromosome N50
Consensus quality (QV)	Primary: 63.6; alternate: 64.5; combined: 64.2	≥ 40
k-mer completeness	Primary: 72.96%; alternate: 72.67%; combined: 98.80%	≥ 95%
BUSCO	C:88.5%[S:87.5%; D:1.0%]; F:4.8%; M:6.7%; n:5 295	single > 90%; duplicated < 5%
Percentage of sequence assigned to chromosomes	99.78%	≥ 90%

1.42 coding transcripts per gene and 5.85 exons per transcript. For further information about the annotation, please refer to the [annotation page](#) on Ensembl.

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: *Patella ulyssiponensis* (China limpet). Accession number [PRJEB63446](#). The genome sequence is released openly for reuse. The *Patella ulyssiponensis* 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. Raw data and assembly accession identifiers are reported in [Table 1](#) and [Table 2](#).

Pipelines used for genome assembly at the WSI Tree of Life are available at <https://pipelines.tol.sanger.ac.uk/pipelines>. [Table 5](#) lists software versions used in this study.

Author information

Members of the Marine Biological Association Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.15333539>.

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

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

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

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

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

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/chhyllp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
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	23.10.0	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
samtools	1.19.2	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.6.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.1a.2	https://github.com/c-zhou/yahs

References

- Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Altschul SF, Gish W, Miller W, *et al.*: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Boaventura D, Hawkins SJ, Teixeira H, *et al.*: **Effects of simulated human exploitation of a key grazer, *Patella vulgata*, on rocky-shore assemblages.** *Mar Ecol Prog Ser.* 2002; **533**: 163–76.
[Publisher Full Text](#)
- Bowman RS: **The morphology of *Patella* spp. juveniles in Britain, and some phylogenetic inferences.** *J Mar Biol Assoc UK.* 1981; **61**(3): 647–66.
[Publisher Full Text](#)
- Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)

- 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](#)
- Christiaens J: **Révision de genre *Patella* (Mollusca, Gastropoda).** *Bull Mus Natl Hist Nat.* 1973; **182**(5): 1305–92.
[Publisher Full Text](#)
- Cooper A, Knights AM, Walker S, *et al.*: **The role of limpets in biodiversity patterns and bioerosion on coastal infrastructure: Insights from Plymouth Breakwater.** *Ecol Eng.* 2025; **213**: 107548.
[Publisher Full Text](#)
- Crowley L, Allen H, Barnes I, *et al.*: **A sampling strategy for genome sequencing the british terrestrial arthropod fauna [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 123.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience.* 2021; **10**(2): giab008.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dautzenberg PH: **Supplément à la faunule malacologique des environs de saint-malo.** *La Feuille des Jeunes Naturalistes.* 1906; (423–425): 39–45, 53–60, 73–77.
- Delany J, Myers AA, McGrath D: **Recruitment, immigration and population structure of two coexisting limpet species in mid-shore tidepools, on the West Coast of Ireland.** *J Exp Mar Biol Ecol.* 1998; **221**(2): 221–30.
[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](#)
- Firth LB, Cluble C, McGrath A, *et al.*: **If you can't beat them join them: Enemy shells as refugia from grazing pressures.** *Oceanogr Mar Biol Annu Rev.* 2023; **61**: 329–62.
[Publisher Full Text](#)
- Firth LB, Crowe TP: **Large-scale coexistence and small-scale segregation of key species on rocky shores.** *Hydrobiologia.* 2008; **614**: 233–41.
[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](#)
- Gaffney PM: **Allozyme evidence for species distinctions among English Channel limpets.** *Mar Biol.* 1980; **57**: 195–201.
- 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): 2896–98.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hawkins SJ, Harkin K: **Preliminary canopy removal experiments in algal dominated communities low on the shore and in the shallow subtidal on the isle of man.** *Bot Mar.* 1985; **28**(6): 223–30.
[Publisher Full Text](#)
- Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species.** *bioRxiv.* 2025.
[Publisher Full Text](#)
- Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jeffreys JG: **British Conchology, Volume III, Atlas v.** London: Van Voorst, 1865.
- 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](#)
- Koufopanou V, Reid DG, Ridgway SA, *et al.*: **A molecular phylogeny of the patellid limpets (Gastropoda: Patellidae) and its implications for the origins of their antitropical distribution.** *Mol Phylogenet Evol.* 1999; **11**(1): 138–56.
[PubMed Abstract](#) | [Publisher 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](#)
- 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, Seppay 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'Connor NE, Crowe TP: **Biodiversity loss and ecosystem functioning: distinguishing between number and identity of species.** *Ecology.* 2005; **86**(7): 1783–96.
[Publisher Full Text](#)
- Powell AWB: **The patellid limpets of the world.** *Indo-Pacific Mollusca.* 1973; **3**(5): 75–99.
[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.*: **Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ribeiro PA, Xavier R, Santos AM, *et al.*: **Reproductive cycles of four species of *Patella* (Mollusca: Gastropoda) on the northern and central Portuguese coast.** *J Mar Biol Assoc UK.* 2009; **89**(6): 1215–21.
[Publisher Full Text](#)
- Sá-Pinto A, Branco MS, Alexandrino PB, *et al.*: **Barriers to gene flow in the marine environment: insights from two common intertidal limpet species of the atlantic and mediterranean.** *PLoS One.* 2012; **7**(12): e50330.
[Publisher Full Text](#)
- Sá-Pinto A, Branco M, Harris DJ, *et al.*: **Phylogeny and phylogeography of the genus *Patella* based on mitochondrial DNA sequence data.** *J Exp Mar Biol Ecol.* 2005; **325**(1): 95–110.
[Publisher Full Text](#)
- Thompson RJ: **Distribution and population dynamics of the limpet *Patella aspera* (Lamarck) in bantry bay.** *J Exp Mar Biol Ecol.* 1979; **40**(2): 115–35.
[Publisher Full Text](#)
- Thompson RJ: **Sexual maturity and sex change in *Patella ulyssiponensis*.** *J Mar Biol Assoc UK.* 1980; **60**: 57–67.
- 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, Krashenninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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](#)
- Weber LI, Hawkins SJ: ***Patella aspera* and *P. ulyssiponensis*: Genetic evidence of speciation in the north-east Atlantic.** *Mar Biol.* 2005; **147**(1): 153–62.
[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 17 September 2025

<https://doi.org/10.21956/wellcomeopenres.27197.r130768>

© 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

This manuscript reports the high quality genome assembly of the limpet *Patella ulyssiponensis*. The work carried out by the authors follows gold standard methodologies and therefore there are no major methodological concerns about the assembly and annotation process. A couple of minor notes that may be added concern the exclusive use of muscle tissue for the generation of RNA-seq libraries, which might have prevented gathering sufficient evidence for the annotation of a number of genes expected to display high tissue specificity, or selectively expressed during embryonal stages. With this regard, BUSCO was only run in a genome mode and therefore it would be useful to also report the same statistics obtained while analysing the outcome of the annotation process in the "Genome annotation report" section.

On a separate minor note, the introduction should clarify the apparently misleading common name of the species (i.e. China limpet), since the actual distribution range does not suggest an Asiatic origin.

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: 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 02 September 2025

<https://doi.org/10.21956/wellcomeopenres.27197.r130771>

© 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, UK

This Genome Note reports the sequencing and assembly of the genome of *Patella ulyssiponensis*, a widely distributed limpet with important ecological implications. The methods and genome assembly are of high quality, making this an important resource for ecological and evolutionary studies. I only have a few minor points:

- *Patella ulyssiponensis* in italics in Keywords.
- Why is there a link in "remains uncertain" in the background?

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: Comparative genomics, evolutionary biology, evo-devo

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 September 2025

<https://doi.org/10.21956/wellcomeopenres.27197.r130769>

© 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, Matsue, Shiman,, Japan

The full genome report has undergone thorough analysis and is considered satisfactory.

The following are minor suggestions that should be addressed:

- The following statement likely requires a citation: In related species, sex change can depend on density; this has yet to be tested in *P. ulyssiponensis*.
- The BUSCO analysis results (in Assembly quality metrics) appear to fall below the EBP benchmark. It would be beneficial to include a statement indicating whether these values are comparable to typical figures for the closely related genus *Patella*.

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: Evolutionary Biology, 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 26 August 2025

<https://doi.org/10.21956/wellcomeopenres.27197.r129506>

© 2025 Guo X. 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.



Ximing Guo 

Department of Marine and Coastal Sciences, Rutgers University New Brunswick, New Brunswick, New Jersey, USA

The article reports a genome for the china limpet *Patella ulyssiponensis*. The rationale for creating the database is clear and strong. Information on the distribution, ecology and biology of the species is helpful. *P. ulyssiponensis* is a member of a basal gastropod group that holds a unique position in gastropod evolution. A few sentences on its evolutionary significance would be helpful.

The approaches and protocols are appropriate and mostly sound. Some of the methods and results are not fully described. It is not clear how the genome is annotated and what the 22k non-coding genes are. Were the RNA-seq data used for gene annotation? Only one organ/tissue (muscle) was used for RNA-seq. How would the absence of other organs affect gene annotation?

The BUSCO score of the reported genome (88.5%) is low. Well-annotated molluscan genomes have BUSCO scores of 95-99%. Could the low BUSCO score be caused by inadequate gene annotation?

I did not find the annotated genome in gff3 or gff formats, and the protein and CDS sequence files. These data files are most useful.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Partly

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

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molluscan genetics and 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.
