DATA NOTE

The genome sequence of the Brown garden snail, *Cornu aspersum* (O.F.Müller, 1774) (Stylommatophora: Helicidae)

[version 1; peer review: 1 approved, 3 approved with reservations]

Olga Sivell[1], Duncan Sivell[1], Angus Davison[2],
Natural History Museum 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]Natural History Museum, London, England, UK
[2]University of Nottingham, Nottingham, England, UK

v1 **First published:** 08 Sep 2025, **10**:487
https://doi.org/10.12688/wellcomeopenres.24859.1
Latest published: 08 Sep 2025, **10**:487
https://doi.org/10.12688/wellcomeopenres.24859.1

Abstract

We present a genome assembly from an individual *Cornu aspersum* (Brown garden snail; Mollusca; Gastropoda; Stylommatophora; Helicidae). The genome sequence has a total length of 2 908.45 megabases. Most of the assembly (98.95%) is scaffolded into 27 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 14.04 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

Cornu aspersum; Brown garden snail; genome sequence; chromosomal; Stylommatophora

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ? ? ?

	1	2	3	4
version 1 08 Sep 2025	✓ view	? view	? view	? view

1. **Dongxiu Xue**, Qingdao Marine Science and Technology Center, Qingdao, China
2. **Peri Bolton**, The Ohio State University Department of Evolution Ecology and Organismal Biology, Columbus, USA
3. **Maitrya Sil**, Jain University, Bengaluru, India
4. **Deborah Power**, Centro de Ciencias do Mar, Universidade do Algarve, Algarve, Portugal

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: Sivell O: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Sivell D**: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Davison A**: Investigation, Resources;

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: Sivell O, Sivell D, Davison A *et al.* **The genome sequence of the Brown garden snail, *Cornu aspersum* (O.F.Müller, 1774) (Stylommatophora: Helicidae) [version 1; peer review: 1 approved, 3 approved with reservations]** Wellcome Open Research 2025, **10**:487 https://doi.org/10.12688/wellcomeopenres.24859.1

First published: 08 Sep 2025, **10**:487 https://doi.org/10.12688/wellcomeopenres.24859.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Spiralia; Lophotrochozoa; Mollusca; Gastropoda; Heterobranchia; Euthyneura; Panpulmonata; Eupulmonata; Stylommatophora; Helicina; Helicoidea; Helicidae; Cornu; *Cornu*; *Cornu aspersum* (O.F.Müller, 1774) (NCBI:txid6535)

Background

Cornu aspersum (O.F. Müller, 1774), also known as the Common Garden Snail, Brown Garden Snail or simply the Garden Snail, belongs to the family Helicidae. Until relatively recently this species was known as *Helix aspera* (Anderson, 2005; ICZN, 2015).

Cornu aspersum is large species with a globular shell 30–35 mm tall and wide, but it can be as small as 20 mm or as large as 40–45 mm. *C. aspersum* has a variable pattern of contrasting paler and darker brown stripes, bands and mottling, sometimes with white flecking. Unlike other large snails found in Britain the shell surface has an irregularly wrinkled texture that is visible on close inspection or under low magnification. Mature snails have a white lip around the mouth of the shell. More details on colour forms, morphology and identification can be found in Taylor (1914), Cameron and Redfern (1976) and Cameron and Riley (2008).

Cornu aspersum is nocturnal but can be active during the day if humidity levels are high (Attia, 2004). This snail prefers to eat living plant material and is a known pest of gardens and horticulture (Cameron, 2016; Stearns, 1900). This species can also be farmed for human and animal consumption (Cowie, 2022; Guiller *et al.*, 2012; Taylor, 1912). *C. aspersum* typically live for 3 to 5 years, are hemaphrodites and lay eggs in the soil in clutches of 40–100 (Cowie, 2022; Madec *et al.*, 2000). *C. aspersum* is known to suffer high rates of winter mortality in colder areas (Boycott, 1934; Taylor, 1914), but populations can rebound following a crash, partly due to an adaptable reproductive strategy (Madec *et al.*, 2000). These snails hibernate for 4-5 months in central England (Boycott, 1934) and for up to 7 months in Scotland (Iglesias *et al.*, 1996). Individual snails can show a strong homing instinct over short distances (e.g. tens of metres) and may repeatedly return to the same resting spot day after day (Brooks, 2011; Dunstan & Hodgson, 2014; Taylor, 1914).

In Britian and Ireland *Cornu aspersum* is common and widespread across southern and lowland areas. Further north it has a more coastal distribution, especially beyond the central belt of Scotland, extending to Orkney and Shetland (Kerney, 1999; NBN Atlas Partnership, 2024). Low winter temperatures are an important factor influencing this species' distribution (Cameron, 2016; Kerney, 1999). In southern Britain on more neutral or base-rich soils *C. aspersum* occurs in a wide variety of habitats such as gardens, parks, dunes, woods and hedgerows (Cameron & Riley, 2008; Kerney, 1999). In northern Britain *C. aspersum* tends to occupy more coastal or synanthropic habitats where it avoids the more acidic soils and areas with colder winter temperatures (Cameron & Redfern, 1976).

Cornu aspersum originates in the Mediterranean region but has been introduced to most continents (Cowie, 2022; GBIF, 2025; Guiller *et al.*, 2012; Taylor, 1912). This snail was likely dispersed by the Romans through their empire to be reared for food (Horšák *et al.*, 2013). Since the 16th century *C. aspersum* has spread to North and South America, South Africa and Oceania (Cowie, 2022), including an intentional introduction from France to California in the 1850s (Stearns, 1900). It has colonised Mauritius and has become established in Hawaii (Cameron, 2016).

The high-quality genome of *Cornu aspersum* was sequenced from a single specimen (NHMUK014449263; SAMEA8534452) from Luton, England. It will aid research on taxonomy and biology of the species and the phylogeny of the family Helicidae. The genome was sequenced as part of the Darwin Tree of Life Project, a collaborative effort to sequence all named eukaryotic species in the Atlantic Archipelago of Britain and Ireland.

Methods

Sample acquisition and DNA barcoding

The specimen used for genome sequencing was an adult *Cornu aspersum* (specimen ID NHMUK014449263, ToLID xgCorAspe16; Figure 1), collected from Luton, England (latitude 51.8882, longitude –0.3851) on 2020-10-20. The specimen was collected by Olga Sivell and Duncan Sivell and formally identified by Duncan Sivell (Natural History Museum). Two other specimens were used for Hi-C sequencing (specimen ID SAN0001335, ToLID xgCorAspe2) and RNA sequencing (specimen ID SAN0001344, ToLID xgCorAspe11). They were collected by Angus Davison from Beeston, Nottingham, United Kingdom (latitude 52.9147, longitude –1.1894) on 2019-12-05. 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

Figure 1. Photograph of the *Cornu aspersum* (xgCorAspe16) specimen used for genome sequencing.

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 xgCorAspe16 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.

RNA was extracted from tissue of xgCorAspe11 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol. The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

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

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

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of tissue of the xgCorAspe2 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

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

RNA library preparation and sequencing

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

Genome assembly

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

The HiFi reads were assembled using Hifiasm (Cheng *et al.*, 2021) with the --primary option. 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 79 breaks, 188 joins, and removal of 31 haplotypic duplications. 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 *Cornu aspersum* specimen generated 115.56 Gb (gigabases) from 11.06 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 2 551.46 Mb, with a heterozygosity of 1.68% and repeat content of 58.88% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 43× coverage. Hi-C sequencing produced 441.30 Gb from 2 922.54 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.

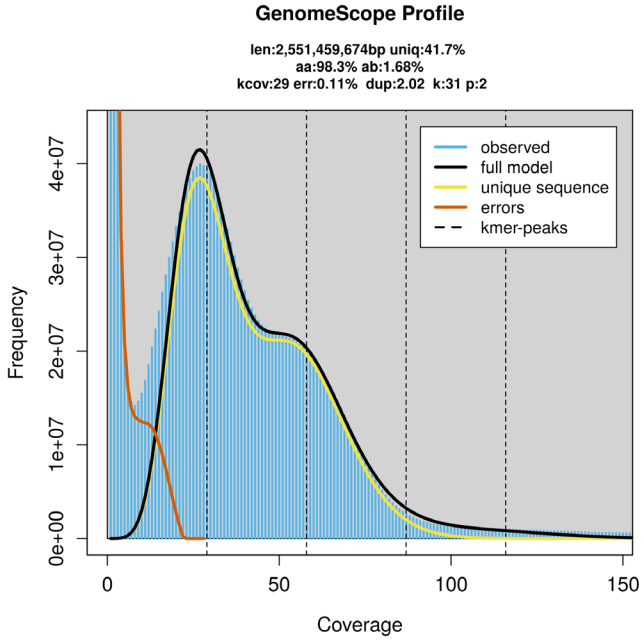

Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

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 2 908.45 Mb in 603 scaffolds, with 3 844 gaps, and a scaffold N50 of 110.3 Mb (Table 2).

Most of the assembly sequence (98.95%) was assigned to 27 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

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

The combined primary and alternate assemblies achieve an estimated QV of 55.5. The *k*-mer completeness is 77.72% for the primary assembly, 62.70% for the alternate haplotype, and 97.89% for the combined assemblies (Figure 4).

BUSCO v.5.7.1 analysis using the mollusca_odb10 reference set (n = 5 295) identified 89.4% of the expected gene set (single = 71.9%, duplicated = 17.5%). 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

Table 1. Specimen and sequencing data for BioProject PRJEB75283.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	xgCorAspe16	xgCorAspe2	xgCorAspe11
Specimen ID	NHMUK014449263	SAN0001335	SAN0001344
BioSample (source individual)	SAMEA8534452	SAMEA9312698	SAMEA9312911
BioSample (tissue)	SAMEA8534533	SAMEA9312776	SAMEA9312975
Tissue	anterior body	**other somatic animal tissue**	**other somatic animal tissue**
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR12954123; ERR12954125; ERR12954120; ERR12954121; ERR12954122; ERR12954124	ERR12982566	ERR12982567
Read count total	11.06 million	2 922.54 million	73.88 million
Base count total	115.56 Gb	441.30 Gb	11.16 Gb

Table 2. Genome assembly statistics.

Assembly name	xgCorAspe16.1
Assembly accession	GCA_964187895.1
Alternate haplotype accession	GCA_964187865.1
Assembly level	chromosome
Span (Mb)	2 908.45
Number of chromosomes	27
Number of contigs	4 447
Contig N50	1.3 Mb
Number of scaffolds	603
Scaffold N50	110.3 Mb
Organelles	Mitochondrion: 14.04 kb

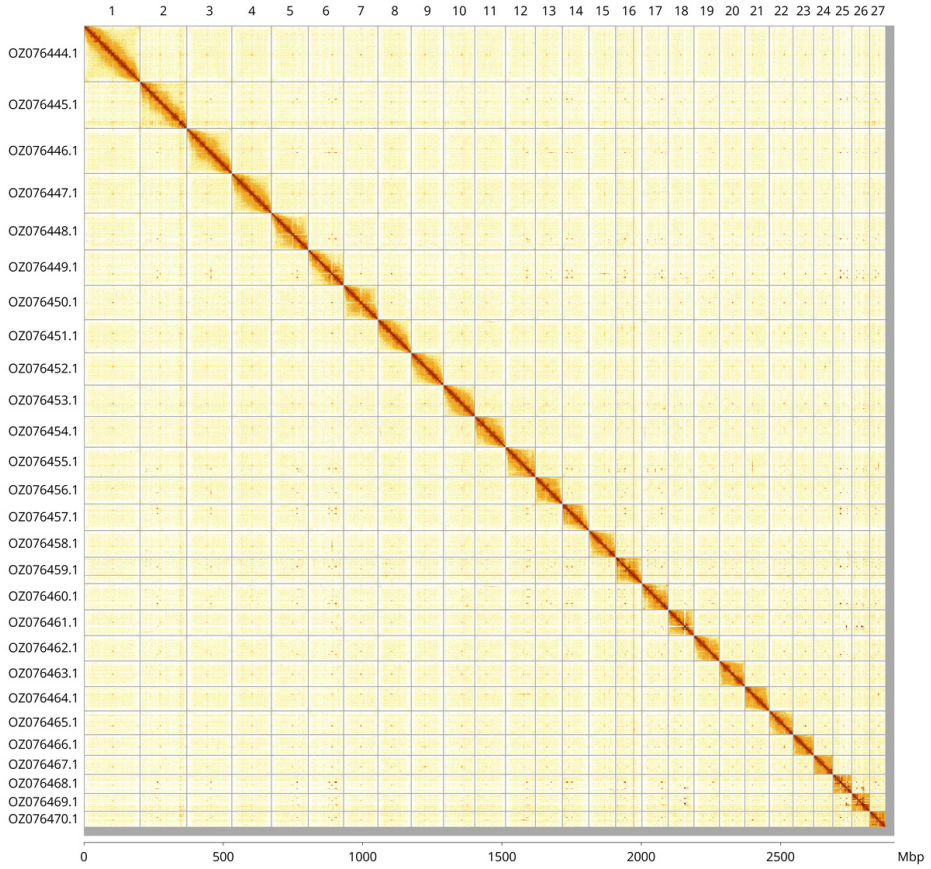

Figure 3. Hi-C contact map of the *Cornu aspersum* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Cornu aspersum* xgCorAspe16.

INSDC accession	Molecule	Length (Mb)	GC%
OZ076444.1	1	201.86	40.50
OZ076445.1	2	167.99	41
OZ076446.1	3	161.73	41
OZ076447.1	4	141.78	41
OZ076448.1	5	132.20	41
OZ076449.1	6	127.11	41
OZ076450.1	7	123.25	41
OZ076451.1	8	119.04	41
OZ076452.1	9	115.96	41
OZ076453.1	10	112.17	41
OZ076454.1	11	110.30	41
OZ076455.1	12	107.26	41
OZ076456.1	13	96.67	41
OZ076457.1	14	95.72	40.50
OZ076458.1	15	95.31	41
OZ076459.1	16	94.90	41
OZ076460.1	17	94.58	40.50
OZ076461.1	18	92.02	40.50
OZ076462.1	19	91.50	41
OZ076463.1	20	91.08	41
OZ076464.1	21	87.65	41
OZ076465.1	22	85.66	41
OZ076466.1	23	73.45	41.50
OZ076467.1	24	68.77	41
OZ076468.1	25	68.32	41
OZ076469.1	26	64.34	40.50
OZ076470.1	27	57.16	41

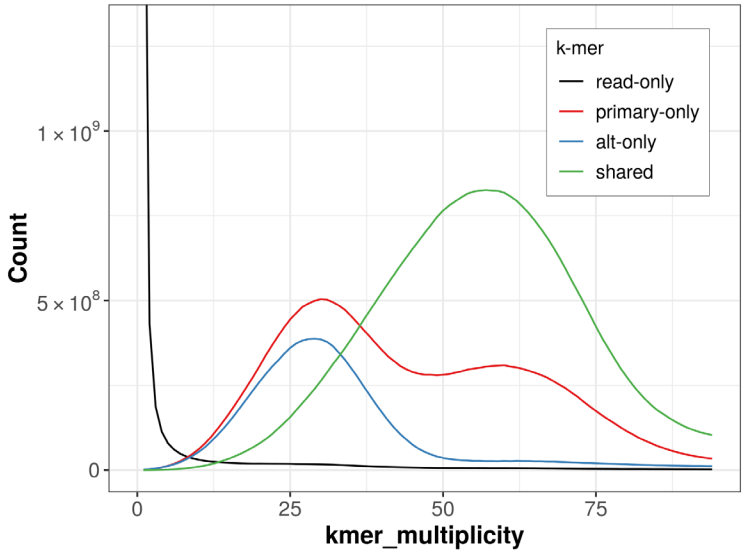

Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

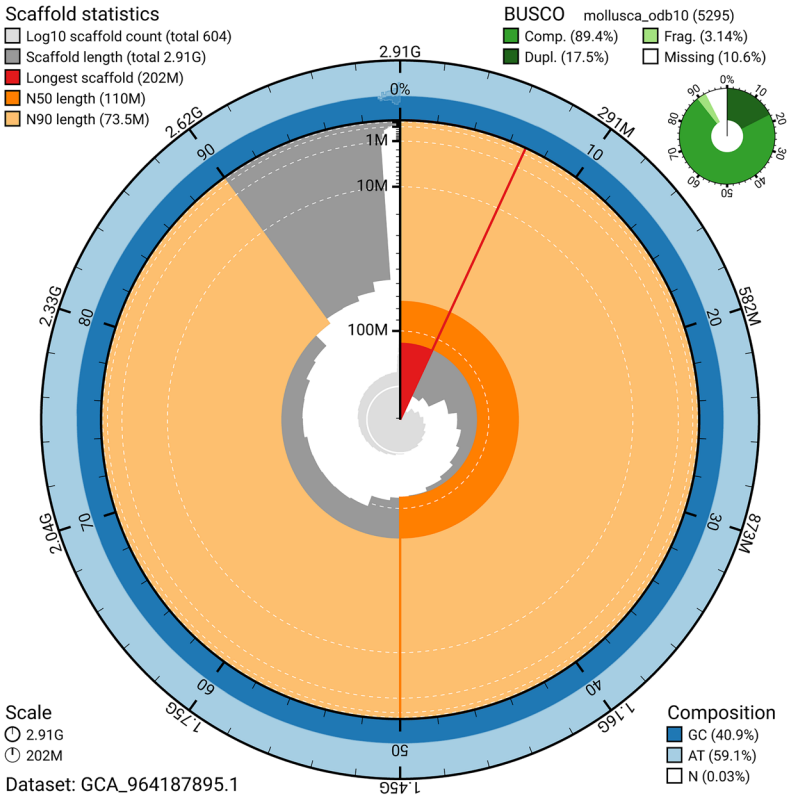

Figure 5. Assembly metrics for xgCorAspe16.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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.

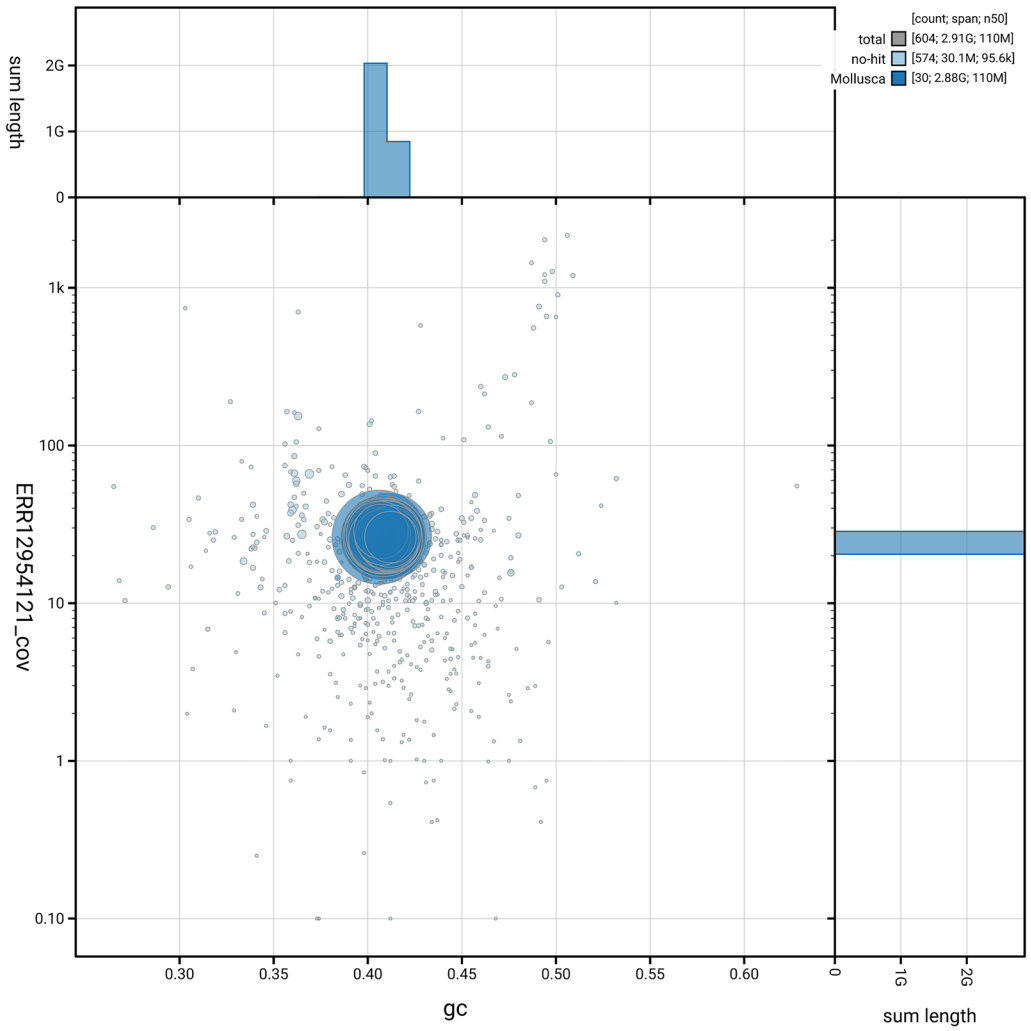

Figure 6. BlobToolKit GC-coverage plot for xgCorAspe16.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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 *Cornu aspersum* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q55	6.C.Q40
Contig N50 length	1.30 Mb	≥ 1 Mb
Scaffold N50 length	110.30 Mb	= chromosome N50
Consensus quality (QV)	Primary: 55.6; alternate: 55.4; combined: 55.5	≥ 40
k-mer completeness	Primary: 77.72%; alternate: 62.70%; combined: 97.89%	≥ 95%
BUSCO	C:89.4% [S:71.9%; D:17.5%]; F:3.1%; M:7.5%; n:5 295	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	98.95%	≥ 90%

Assembly Standards September 2024. The EBP metric, calculated for the primary assembly, is **6.C.Q55**, meeting the recommended reference standard.

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: Cornu aspersum (brown garden snail). Accession number PRJEB75283. The genome sequence is released openly for reuse. The *Cornu aspersum* 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 Natural History Museum 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	4.4.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.7.1	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

Software	Version	Source
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
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
Minimap2	2.28-r1209	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	24.10.4	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	1.21	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.1	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

Anderson R: **An annotated list of the non-marine mollusca of Britain and Ireland.** *J Conchol.* 2005; **38**(6): 607–37.
Publisher Full Text

Attia J: **Behavioural rhythms of land snails in the field.** *Biol Rhythm Res.* 2004; **35**(1–2): 35–41.
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

Boycott AE: **The habitats of land mollusca in Britain.** *J Ecol.* 1934; **22**(1): 1–38.
Publisher Full Text

Brooks R: **The homing instinct of *Cornu aspersum*.** *Mollusc World.* 2011; **26**: 4–6.

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

Cameron R: **Slugs and snails**. William Collins, 2016.
Reference Source

Cameron R, Riley G: **Keys for the identification of land snails of the British Isles**. Telford: Field Studies Council, 2008.
Reference Source

Cameron RAD, Redfern M: **British land snails**. London: Academic Press, 1976.
Reference Source

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

Cowie R: ***Cornu aspersum* (common garden snail)**. 2022.
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

Dunstan DJ, Hodgson DJ: **Snails home.** *Phys Scr.* 2014; **89**: 068002.
Publisher Full Text

Ewels P, Magnusson M, Lundin S, *et al.*: **MultiQC: summarize analysis results for multiple tools and samples in a single report.** *Bioinformatics.* 2016; **32**(19): 3047–3048.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ewels PA, Peltzer A, Fillinger S, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines.** *Nat Biotechnol.* 2020; **38**(3): 276–278.
PubMed Abstract | **Publisher Full Text**

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

GBIF: ***Cornu aspersum* (O. F. Müller, 1774) map on GBIF**. 2025.
Reference Source

Guiller A, Martin MC, Hiraux C, *et al.*: **Tracing the invasion of the Mediterranean land snail *Cornu aspersum aspersum* becoming an agricultural and garden pest in areas recently introduced.** *PLoS One.* 2012; **7**(12): e49674.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Horšák M, Juřičková L, Picka J: **Molluscs of the Czech and Slovak Republics**. Zlín: Kabourek Press, 2013.
Reference Source

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**

ICZN: **Opinion 2354 (Case 3518). *Cornu* Born, 1778 (Mollusca, Gastropoda, Pulmonata, Helicidae): request for a ruling on the availability of the generic name granted.** *Bull Zool Nomencl.* 2015; **72**(2): 157–58.
Publisher Full Text

Iglesias J, Santos M, Castillejo J: **Annual activity patterns of the land snail *Helix aspersa* Müller in natural populations in north-western Spain.** *J Molluscan Stud.* 1996; **62**: 495–505.
Publisher Full Text

Kerney M: **Atlas of land and freshwater molluscs of Britain and Ireland**. Colchester: Harley Books, 1999.
Reference Source

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**

Madec L, Desbuquois C, Coutellec-Vreto MA: **Phenotypic plasticity in reproductive traits: importance in the life history of *Helix aspersa* (Mollusca: Helicidae) in a recently colonized habitat.** *Biol J Linn Soc.* 2000; **69**(1): 25–39.
Publisher 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

NBN Atlas Partnership: ***Cornu aspersum* (O. F. Müller, 1774) map on the NBN Atlas.** 2024.
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**

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**

Stearns REC: **Exotic Mollusca in California.** *Science.* 1900; **11**: 655–59.
Publisher Full Text

Taylor JW: **Biology of the mollusca: based chiefly upon a study of one of our commonest species *Helix aspersa.*** *J Conchol.* 1912; **13**(9): 265–73.
Publisher Full Text

Taylor JW: **Monograph of the land and freshwater Mollusca of the British Isles, zonitidae, endodontidae, helicidae**. Leeds, UK: Taylor Brothers, 1914.
Reference Source

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 26 December 2025

https://doi.org/10.21956/wellcomeopenres.27383.r140155

?

Deborah Power
Campus de Gambelas, Centro de Ciencias do Mar, Universidade do Algarve, Algarve, Portugal

I enjoyed reading this MS as it is clearly written and concise. The objective is clear and so my evaluation is primarily linked to the fundamental aim, which is to generate high quality genomes and provide open acess to the data and the means of their generation. As such I consider that overall this has been acheived particulalry since much of the methodology of the study complies with the protocols.io developed as part of the "Tree of life" exercise. However, there were some aspects, related to parts of the methods that I feel could be improved to make the technical approach totally transparent, and reproducible. I make some observations about additional information that would be useful to have in the Introduction and then some small comments related to parts of the method description. I briefly identify my main suggestions/comments below:
1. In the introduction it would have been helpful to have an indication that the *Cornu aspersum* is diploid.
2. It is indicated that *C. aspersum* is hemaphrodite but it would have been useful to indicate that they undergo sexual reproduction generally.
3. Linked to the previous point I found the comment that they have an "adaptable reproductive strategy" without real meaning, and perhaps here a short but specific statement about their reproductive strategy could be provided.
4. I had a question about if/how using specimens from different geographical locations for the genome and transcriptomes influenced the results.
5. I guess the mitochondrial DNA was only obtained for the specimen from Luton. That might have given some insight into the effect of geography or perhaps this would have been found if the transcriptome was compared to the genome.
6. In the section about seqeuncing, I could not find reference to protocols.io for the sequencing approach and so they seem very vague compared to the more precise methodology reports of the other sections. Hi-C I would have apprecated a it more definition in relation to the reagents e.g., Arima Genomics, but country not mentioned, Diagnocine Power Masher-II (no fabricant mentioned), Qubit HS Assay Kit - is this the same as the Qubit dsDNA high sensitivity assay??, its not clear as written, Covaris E220 sonicator is from where and what were the setting used?
I could continue in relation to the library preparation, but I hope the points I raise indicate what I

think is missing in general from this section, which as I said does not seem to be covered by protocols.io and so makes this section of the MS much less clear.
Assembly methods are thorough and comprehensible.
One comment about the Genome sequence report. I suggest a change to the following statement, I find it ambiguous and imprecise:
These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size....
This I think would be clearer if written as follows:
These chromosome-level scaffolds, confirmed by Hi-C data, are sequentially numbered from largest to smallest size....
In Table 1, why is "other somatic animal tissue" indicated under tissue. This is ambiguous particulalry since there virtually no description about what tissues were harvested and how. Similarly, it is indicated anterior body was used for the genome sequencing but not explained anywhere in the methods how this is identified anatomically.

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?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Comparative Endocrinology, Comparative Genomics, Fish, Bivalves, Invertebrates, Biotechology

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 13 November 2025

https://doi.org/10.21956/wellcomeopenres.27383.r134586

?

Maitrya Sil

Jain University, Bengaluru, Karnataka, India

This is a well written manuscript. The quality of the assembly looks good. The genome completeness is a little low. The authors should provide reasons behind the low completeness. I suggest using the **most recent version of BUSCO** to alleviate this issue. The accession number of the mitogenome is also missing. If has been submitted separately, I suggest that the authors provide the accession number.

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

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

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

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

Competing Interests: No competing interests were disclosed.

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

Reviewer Report 13 October 2025

https://doi.org/10.21956/wellcomeopenres.27383.r134584

?

Peri Bolton
The Ohio State University Department of Evolution Ecology and Organismal Biology, Columbus, Ohio, USA

Generally good standard long-read genome assembly paper, just some extra thoughts:

The paragraph on the distribution feels a little bit repetitive in discussion of coastal distribution.

One thing I am not sure I understand: The GC content on the snail plot is basically completely uniform across the genome. I was under the impression that GC content varied across the genome, so I do not understand why this plot does not reflect this. However, I see that many other assemblies feature this, so I am not sure that this is exactly a red flag.

This is in response to my "Partly": Not sure why the genome annotation wasn't presented in this manuscript, given that RNAseq was conducted and reported. The authors suggest that it will be done in the future. I do not know if this is standard practice for these genome reports. I have found examples of others that do include annotation reports.

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?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Conservation 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 09 October 2025

https://doi.org/10.21956/wellcomeopenres.27383.r134588

Dongxiu Xue
Qingdao Marine Science and Technology Center, Qingdao, China

This manuscript presents a high-quality, chromosome-level genome assembly of *Cornu aspersum* generated under the Darwin Tree of Life Project. The use of PacBio HiFi and Hi-C data has produced a robust assembly and provides an important reference for terrestrial gastropods. The detailed documentation of protocols and data accessibility is exemplary.
however, the authors should clarify whether the individual used for PacBio HiFi sequencing is of the same sex as those used for transcriptome annotation and Hi-C scaffolding. If the sex of each specimen is known, it would be useful to state this explicitly in the text, as potential sex-linked genomic differences or sex-biased expression may influence assembly structure and annotation outcomes.

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, genome evolution, population 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.