



DATA NOTE

The chromosomal genome sequence of the buttercup lucine, *Anodontia alba* Link, 1807 (Lucinida: Lucinidae) and its associated microbial metagenome sequences

[version 1; peer review: 2 approved]

Diana Chin ¹, Barbara Campbell ², Jillian Petersen³, Shen Jean Lim ⁴, Graeme Oatley ⁵, Elizabeth Sinclair ⁵, Eerik Aunin⁵, Noah Gettle ⁵, Camilla Santos⁵, Michael Paulini⁵, Haoyu Niu⁵, Victoria McKenna⁵, Rebecca O'Brien ⁵,

Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,

Wellcome Sanger Institute Scientific Operations: Sequencing Operations,

Wellcome Sanger Institute Tree of Life Core Informatics team,

EBI Aquatic Symbiosis Genomics Data Portal Team,

Aquatic Symbiosis Genomics Project Leadership

¹Marine Science Center, Northeastern University, Nahant, MA, USA

²Department of Biological Sciences, Clemson University, Clemson, SC, USA

³Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria

⁴College of Marine Science, University of South Florida, St. Petersburg, FL, USA

⁵Tree of Life Programme, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 17 Feb 2026, 11:131
<https://doi.org/10.12688/wellcomeopenres.25920.1>
Latest published: 17 Feb 2026, 11:131
<https://doi.org/10.12688/wellcomeopenres.25920.1>

Abstract

We present a genome assembly from an individual *Anodontia alba* (buttercup lucine; Mollusca; Bivalvia; Lucinida; Lucinidae). The genome sequence has a total length of 1 862.85 megabases. Most of the assembly (99.28%) is scaffolded into 18 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 18.48 kilobases. Gene annotation of this assembly by Ensembl identified 12 083 protein-coding genes. From the metagenome data, we recovered four bins, of which three were high-quality MAGs.

Keywords

Anodontia alba; buttercup lucine; genome sequence; chromosomal; Lucinida; microbial metagenome assembly

Open Peer Review

Approval Status

	1	2
version 1		
17 Feb 2026	view	view
1. Mohamed Abdelsalam , Cairo University, Giza, Egypt		
2. Craig Wilding , University College Dublin, Dublin, Ireland		
Any reports and responses or comments on the article can be found at the end of the article.		



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

Corresponding author: Aquatic Symbiosis Genomics Project Leadership (mark.blaxter@sanger.ac.uk)

Author roles: **Chin D:** Investigation, Resources; **Campbell B:** Investigation, Resources, Writing – Review & Editing; **Petersen J:** Conceptualization, Project Administration, Writing – Review & Editing; **Lim SJ:** Investigation, Resources, Writing – Original Draft Preparation; **Oatley G:** Investigation; **Sinclair E:** Investigation; **Aunin E:** Formal Analysis; **Gettle N:** Formal Analysis; **Santos C:** Formal Analysis; **Paulini M:** Formal Analysis; **Niu H:** Project Administration; **McKenna V:** Project Administration; **O'Brien R:** Project Administration;

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by the Gordon and Betty Moore Foundation through a grant (GBMF8897) to the Wellcome Sanger Institute to support the Aquatic Symbiosis Genomics Project, and by Wellcome through core funding to the Wellcome Sanger Institute (220540). JMP was funded by an ERC Starting Grant EvoLucin and a WWTF Vienna Research Group grant.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Chin D *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: Chin D, Campbell B, Petersen J *et al.* **The chromosomal genome sequence of the buttercup lucine, *Anodontia alba* Link, 1807 (Lucinida: Lucinidae) and its associated microbial metagenome sequences [version 1; peer review: 2 approved]** Wellcome Open Research 2026, 11:131 <https://doi.org/10.12688/wellcomeopenres.25920.1>

First published: 17 Feb 2026, 11:131 <https://doi.org/10.12688/wellcomeopenres.25920.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Spiralia; Lophotrochozoa; Mollusca; Bivalvia; Autobranchia; Heteroconchia; Euheterodonta; Imparidentia; Lucinida; Lucinoidea; Lucinidae; *Anodontia*; *Anodontia alba* Link, 1807 (NCBI:txid244458)

Background

Anodontia alba, also known as the buttercup lucine, was first described by Link in 1807 (Link, 1807). It belongs to the species-rich and widely distributed family of chemosymbiotic marine bivalves called Lucinidae. Lucinidae have a fossil record dating back to the Silurian (Liljedahl, 1992), and fossils similar to extant *Anodontia* have been reported (Taylor & Glover, 2005). *Anodontia alba* is placed within the morphologically diverse Leucosphaerinae subfamily (Taylor et al., 2016). All extant Lucinidae species house chemosynthetic intracellular bacteria within their gill epithelial cells (Taylor & Glover, 2021). Known lucinid symbionts fix inorganic carbon using the energy gained from oxidising reduced sulfur compounds. This carbon is a source of nutrition for the symbionts and for their hosts. The endosymbionts can also significantly reduce the amount of sulfide in the environments they colonise, including root zones of seagrasses and seagrass sediments to make the environment less toxic (Chin et al., 2021; Reynolds et al., 2007; van der Heide et al., 2012). In addition to their role in nutrient cycling, lucinids are prey for commercially important species such as lobsters (Higgs et al., 2016).

Anodontia alba is typically found in coastal seagrass beds and is distributed in the western Atlantic, including Florida (Lim et al., 2019), Guadeloupe (Taylor et al., 2016), southern Gulf of Mexico (Suárez-Mozo et al., 2023), and as far south as Panama in the Caribbean Sea (Osvatic et al., 2021). It can grow up to 6 cm in size, and its shell has a white, smooth, sub-spherical, and toothless appearance (Taylor et al., 2016). Like many lucinid symbionts found in seagrass sediments (König et al., 2016; Morel-Letelier et al., 2024; Osvatic et al., 2023; Petersen et al., 2016), the *Anodontia alba* symbiont from Panama has the genetic capability to fix nitrogen in addition to carbon (Osvatic et al., 2021).

Chemosynthetic symbiosis evolved several times independently in diverse animal groups (Sogin et al., 2021), and has resulted in major behavioural, physiological, and morphological adaptations in lucinids (Taylor & Glover, 2021), which are presumably genetically encoded. However, although genome sequences are available for several other bivalve and non-bivalve hosts of chemosynthetic symbionts, few Lucinidae genomes have been sequenced (Guo et al., 2023; e.g. Sun et al., 2017). Genomes of *Luciniscia nassula* (Anderson et al., 2025), *Ctena decussata* (Wilkins et al., 2025) and *Indoaustriella scarlatoi* (Guo et al., 2025) have recently been published. The genome of *A. alba* presented here will enable a new understanding of how chemosymbiosis has shaped the evolution of lucinid bivalves and allow us to identify shared or unique genomic features across animal groups that are relevant to symbiosis and/or chemosynthetic symbiosis.

Methods

Sample acquisition

An *Anodontia alba* specimen (specimen ID AG1/VIEM1130001, ToLID xbAnoAlba1; Figure 1) was collected from Sammy Creek Landing, FL, USA on 2022-02-06, where this species was previously found co-existing with at least five other lucinid species (Lim et al., 2019). The specimen was taken from its habitat of seagrass sediments by Barbara Campbell (Clemson University) and Shen Jean Lim (University of South Florida) using a shovel and sieve. The specimen was identified by Diana Chin (Northeastern University) based on gross morphology. Dissected tissues were placed immediately in a microcentrifuge tube and flash-frozen in a mixture of dry ice and ethanol. The same specimen was used for RNA sequencing.

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 xbAnoAlba1 sample was weighed and triaged to determine the appropriate extraction protocol. Gill tissue was homogenised by powermashing using a PowerMasher II tissue disruptor. HMW DNA was extracted using the Manual Tissue Nanobind 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 Femto Pulse system. For this sample, the final post-shearing DNA had a Qubit concentration of 159.0 ng/μL and a yield of 11 925.00 ng. The Femto Pulse Genomic Quality Number (GQN) was 5.40 (threshold 10 kb).

RNA was also extracted from tissue of xbAnoAlba1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ mirVana protocol. The RNA concentration



Figure 1. Photographs of the *Anodontia alba* (xbAnoAlba1) specimen used for genome sequencing.

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 frozen gill tissue from the xbAnoAlba1 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 (ThermoFisher 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 to create 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, generating 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 OATK (Zhou *et al.*, 2025).

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 53 breaks, 61 joins, and removal of 59 haplotypic duplications, which reduced the scaffold count by 18.8%, increased the scaffold N50 by 4.2%, and reduced the total assembly length by 1.0%. The curation process is described at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextView Snapshot 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](#)).

Metagenome assembly

The metagenome assembly was generated using MetaMDBG ([Benoit et al., 2024](#)) and binned using MetaBAT2

([Kang et al., 2019](#)). The resulting bin sets of each binning algorithm were optimised and refined using DAS Tool ([Sieber et al., 2018](#)). PROKKA ([Seemann, 2014](#)) was used to identify tRNAs and rRNAs in each bin, CheckM ([Parks et al., 2015](#)) (checkM_DB release 2015-01-16) was used to assess bin completeness/contamination, and GTDB-Tk ([Chaumeil et al., 2022](#)) (GTDB release 214) was used to taxonomically classify bins. Taxonomic replicate bins were identified using dRep ([Olm et al., 2017](#)) with default settings (95% ANI threshold). All bins were assessed for quality and categorised as metagenome-assembled genomes (MAGs) if they met the following criteria: contamination $\leq 5\%$, presence of 5S, 16S, and 23S rRNA genes, at least 18 unique tRNAs, and either $\geq 90\%$ completeness or $\geq 50\%$ completeness with fully circularised chromosomes ([Bowers et al., 2017](#)). Bins that did not meet these thresholds, or were identified as taxonomic replicates of MAGs, were retained as 'binned metagenomes' provided they had $\geq 50\%$ completeness and $\leq 10\%$ contamination.

Genome sequence report

Sequence data

PacBio sequencing of the *Anodontia alba* specimen generated 159.31 Gb (gigabases) from 22.06 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 2 171.11 Mb, with a heterozygosity of 1.99% and repeat content of 55.18% ([Figure 2](#)). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 50 $\times$ coverage. Hi-C sequencing

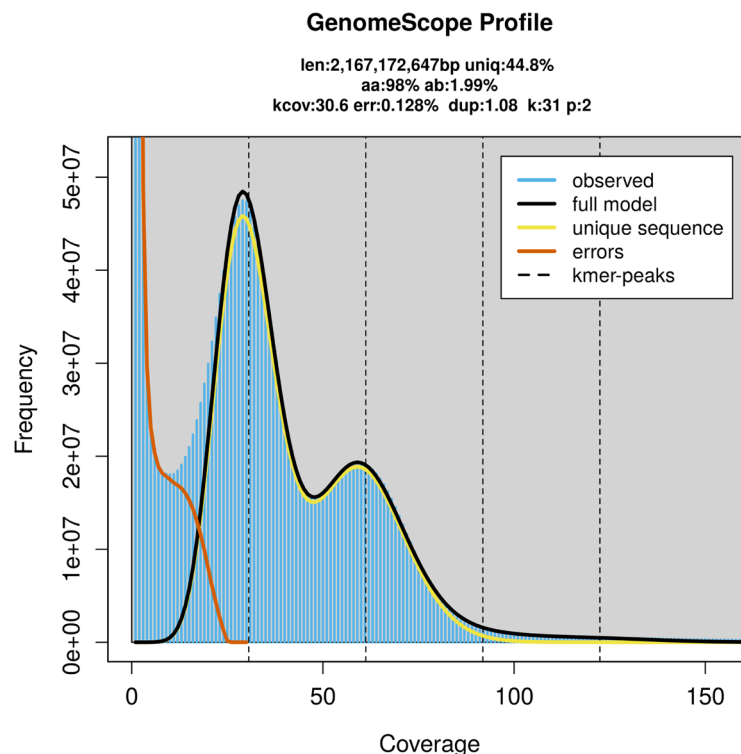


Figure 2. Frequency distribution of k -mers generated using GenomeScope2. The plot shows observed and modelled k -mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

produced 199.50 Gb from 1 321.21 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 1 862.85 Mb in 336 scaffolds, with 1 628 gaps, and a scaffold N50 of 114.1 Mb ([Table 2](#)).

Most of the assembly sequence (99.28%) was assigned to 18 chromosomal-level scaffolds. These chromosome-level

scaffolds, confirmed by Hi-C data, are named according to size ([Figure 3](#); [Table 3](#)).

The mitochondrial genome was also assembled (length 18.48 kb, OZ024839.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

The combined primary and alternate assemblies achieve an estimated QV of 56.2. The *k*-mer completeness is 72.36% for the primary assembly, 72.27% for the alternate haplotype, and 98.20% for the combined assemblies ([Figure 4](#)).

Table 1. Specimen and sequencing data for BioProject PRJEB73659.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	xbAnoAlba1	xbAnoAlba1	xbAnoAlba1
Specimen ID	VIEM1130001	VIEM1130001	VIEM1130001
BioSample (source individual)	SAMEA110043978	SAMEA110043978	SAMEA110043978
BioSample (tissue)	SAMEA14431334	SAMEA14431338	SAMEA14431343
Tissue	gill	gill	other somatic animal tissue
Instrument	Sequel IIe	Illumina NovaSeq 6000	Illumina NovaSeq 6000
Run accessions	ERR12736885; ERR12736886; ERR12736882; ERR12736883; ERR12736884	ERR12737279	ERR12737280
Read count total	22.06 million	1 321.21 million	59.40 million
Base count total	159.31 Gb	199.50 Gb	8.97 Gb

Table 2. Genome assembly statistics.

Assembly name	xbAnoAlba1.1
Assembly accession	GCA_964016985.1
Alternate haplotype accession	GCA_964016975.1
Assembly level	chromosome
Span (Mb)	1 862.85
Number of chromosomes	18
Number of contigs	1 964
Contig N50	1.84 Mb
Number of scaffolds	336
Scaffold N50	114.1 Mb
Organelles	Mitochondrion: 18.48 kb

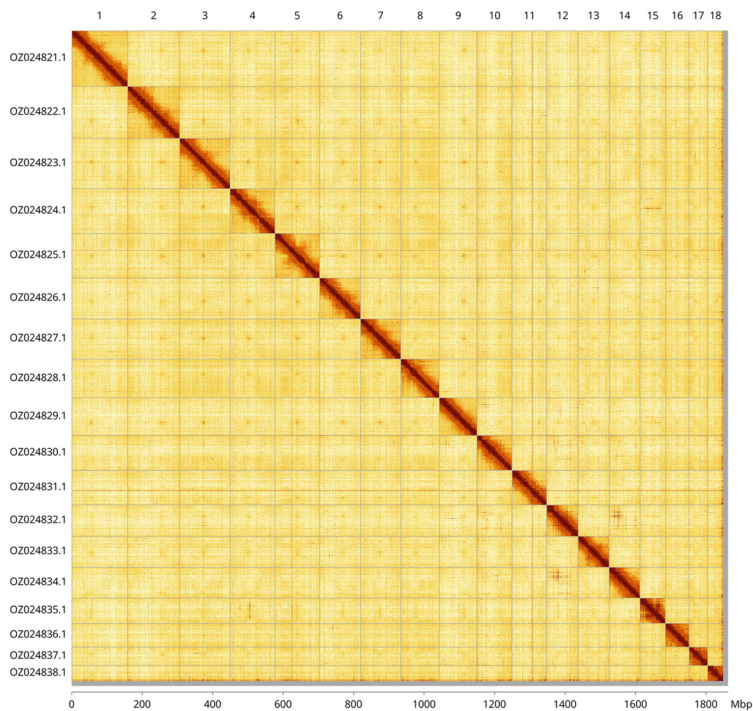


Figure 3. Hi-C contact map of the *Anodontia alba* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Anodontia alba* xbAnoAlba1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ024821.1	1	158.84	37
OZ024822.1	2	147.77	37
OZ024823.1	3	143.40	37
OZ024824.1	4	126.89	37
OZ024825.1	5	126.84	37
OZ024826.1	6	116.18	37.50
OZ024827.1	7	114.10	37
OZ024828.1	8	109.66	37
OZ024829.1	9	106.94	37.50
OZ024830.1	10	99.37	37
OZ024831.1	11	98.49	37
OZ024832.1	12	88.72	37.50
OZ024833.1	13	88.13	37.50
OZ024834.1	14	87.86	37
OZ024835.1	15	71.63	37.50

INSDC accession	Molecule	Length (Mb)	GC%
OZ024836.1	16	67.47	37.50
OZ024837.1	17	52.73	37
OZ024838.1	18	44.36	37.50

BUSCO v.5.5.0 analysis using the mollusca_odb10 reference set ($n = 5\,295$) identified 83.8% of the expected gene set (single = 82.2%, duplicated = 1.5%). 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. \(2021\)](#) and the Earth BioGenome Project Report on Assembly Standards [September 2024](#). The EBP metric, calculated for the primary assembly, is **6.C.Q55**, meeting the recommended reference standard.

Genome annotation report
The *Anodontia alba* genome assembly (GCA_964016985.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 31 170 transcribed mRNAs from 12 083 protein-coding and 11 014 non-coding

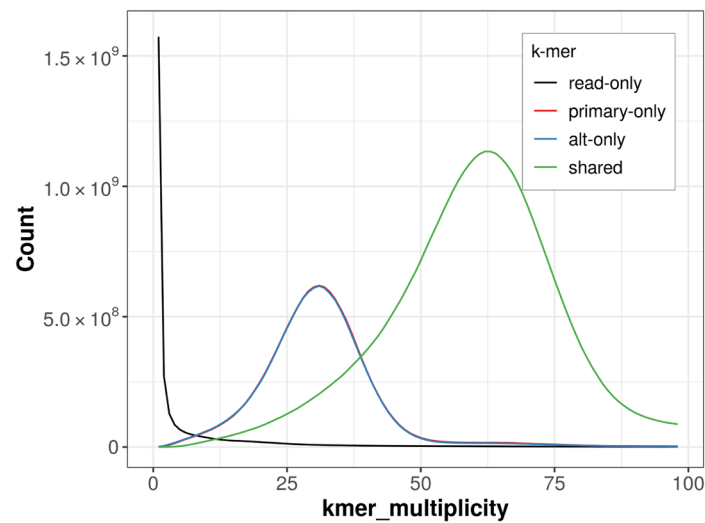


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

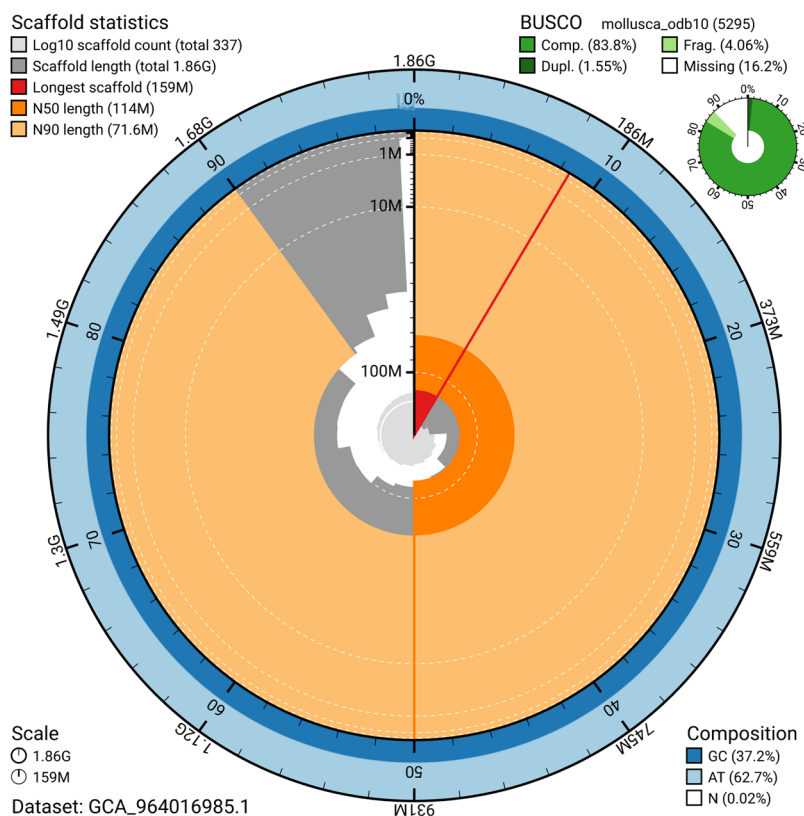


Figure 5. Assembly metrics for xbAnoAlba1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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](#).

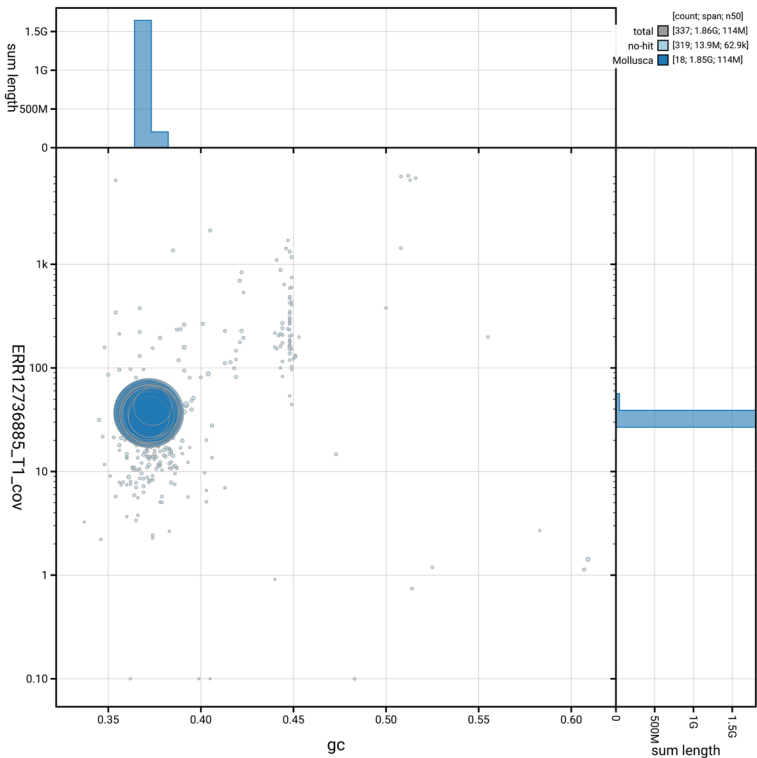


Figure 6. BlobToolKit GC-coverage plot for xbAnoAlba1.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display 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 *Anodontia alba* assembly.

Measure	Value	Benchmark
EBP summary (primary)	6.C.Q55	6.C.Q40
Contig N50 length	1.84 Mb	≥ 1 Mb
Scaffold N50 length	114.10 Mb	= chromosome N50
Consensus quality (QV)	Primary: 55.6; alternate: 56.8; combined: 56.2	≥ 40
<i>k</i> -mer completeness	Primary: 72.36%; alternate: 72.27%; combined: 98.20%	≥ 95%
BUSCO	C:83.8% [S:82.2%; D:1.5%]; F:4.1%; M:12.2%; n:5 295	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	99.28%	≥ 90%

Notes: EBP summary uses log10(Contig N50); chromosome-level (C) or log10(Scaffold N50); Q (Merqury QV). BUSCO: C=complete; S=single-copy; D=duplicated; F=fragmented; M=missing; n=orthologues

genes. The average transcript length is 19 295.35 bp, with an average of 1.35 coding transcripts per gene and 5.33 exons per transcript. For further information about the annotation, please refer to the [Ensembl annotation page](#).

Metagenome report

We recovered four bins from the metagenome assembly, of which three met the criteria for MAGs, including one fully circularised genomes. The recovered bins represented 2 bacterial phyla, with genome sizes ranging from 3.32 to 6.66 Mbp

(mean: 4.88 ± 1.21 Mbp). Mean completeness was 90.0% (± 16.1%) with 0.6% (± 0.7%) contamination. [Table 5](#) summarises the taxa and quality of the metagenome bins.

Software tool versions and sources are given in [Table 6](#).

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Tree of Life collaborator. The Wellcome Sanger Institute employs a process whereby due diligence is

Table 5. Quality metrics and taxonomic assignments of the binned metagenomes.

NCBI taxon	Taxid	GTDB taxonomy	Quality	Size (bp)	Contigs	Circular	Mean coverage	Completeness (%)	Contamination (%)
<i>Pararhizobium</i> sp.	1977563	Alphaproteobacteria	Medium	3 321 067	136	No	2.71	62.16	0.15
<i>Candidatus</i> Thiodiazotropha sp.	2847306	Gammaproteobacteria	High	4 408 143	1	Yes	8 128.05	99.39	1.76
<i>Rhodococcus qingshengii</i>	334542	Actinomycetes	High	6 655 471	9	No	8.13	99.18	0
<i>Mycobacterium</i> sp.	1785	Actinomycetes	High	5 155 284	2	No	18.07	99.17	0.40

Table 6. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
bin3C	0.3.3	https://github.com/cerebis/bin3C
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
checkM	2015-01-16	https://ecogenomics.github.io/CheckM/
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
dRep	3.4.0	https://github.com/MrOlm/drep
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
GTDB-Tk	1.2.1	https://github.com/ECogenomics/GTDBTk
Hifiasm	0.19.8-r603	https://github.com/chhyllp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MAGScoT	1.0.0	https://github.com/ikmb/MAGScoT
MaxBin	2.2.7	https://sourceforge.net/projects/maxbin/
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
MetaBAT2	2.15-15-gd6ea400	https://bitbucket.org/berkeleylab/metabat
metaMDBG	Pre-release	https://github.com/GaetanBenoitDev/metaMDBG
metaTOR	Pre-release	https://github.com/koszulab/metaTOR
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.04.1	https://github.com/nextflow-io/nextflow

Software	Version	Source
Oatk	0	https://github.com/c-zhou/oatk
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
Prokka	1.14.5	https://github.com/tseemann/prokka
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.4.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.1a.2	https://github.com/c-zhou/yahs

carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible.

The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Tree of Life collaborator, Genome Research Limited (operating as the Wellcome Sanger Institute) and in some circumstances other Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Anodontia alba*. Accession number [PRJEB73659](#). The genome sequence is released openly for reuse. The *Anodontia alba* genome sequencing initiative is part of the Aquatic Symbiosis Genomics Project (PRJEB43743) 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](#).

Production code used in genome assembly at the WSI Tree of Life is available at <https://github.com/sanger-tol>. [Table 6](#) lists software versions used in this study.

Author information

Contributors are listed at the following links:

- 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 [EBI Aquatic Symbiosis Genomics Data Portal Team](#)
- The [Aquatic Symbiosis Genomics Project leadership](#)

Acknowledgements

Sample collection was permitted by the Florida Fish and Wildlife Conservation Commission through a Special Activity License (SAL-21-2317-SR).

References

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 LC, Campbell B, Long-Fox B, et al.: **The genome sequences of the woven lucine, *Lucinisa nassula* (Conrad, 1846) and a bacterial endosymbiont [version 1; peer review: 1 approved, 1 approved with**

reservations]. *Wellcome Open Res.* 2025; **10**: 225.

[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](#)

Benoit G, Raguideau S, James R, *et al.*: **High-quality metagenome assembly from long accurate reads with metaDBG.** *Nat Biotechnol.* 2024; **42**(9): 1378–1383.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Bowers RM, Kyrpides NC, Stepanauskas R, *et al.*: **Minimum Information about a Single Amplified Genome (MISAG) and a Metagenome-Assembled Genome (MIMAG) of bacteria and archaea.** *Nat Biotechnol.* 2017; **35**(8): 725–731.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Chin DW, de Fouw J, van der Heide T, *et al.*: **Facilitation of a tropical seagrass by a chemosymbiotic bivalve increases with environmental stress.** *J Ecol.* 2021; **109**(1): 204–217.

[Publisher 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](#)

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](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Guo Y, Meng L, Wang M, *et al.*: **Hologenome analysis reveals independent evolution to chemosymbiosis by deep-sea bivalves.** *BMC Biol.* 2023; **21**(1): 51.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Guo Y, Zhong Z, Zhang N, *et al.*: **Chromosome-level genome assembly of the intertidal lucinid clam *Indoaustriella scarlati*.** *Sci Data.* 2025; **12**(1): 275.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Higgs ND, Newton J, Attrill MJ: **Caribbean spiny lobster fishery is underpinned by trophic subsidies from chemosynthetic primary production.** *Curr Biol.* 2016; **26**(24): 3393–3398.

[PubMed Abstract](#) | [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): giab153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

König S, Gros O, Heiden SE, *et al.*: **Nitrogen fixation in a chemoautotrophic lucinid symbiosis.** *Nat Microbiol.* 2016; **2**(1): 16193.

[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](#)

Liljedahl L: **The Silurian *Ilionia prisca*, oldest known deep-burrowing suspension-feeding bivalve.** *J Paleontol.* 1992; **66**(2): 206–210.

[Publisher Full Text](#)

Lim SJ, Alexander L, Engel AS, *et al.*: **Extensive thioautotrophic gill endosymbiont diversity within a single *Ctena orbiculata* (Bivalvia: Lucinidae) population and implications for defining host-symbiont specificity and species recognition.** *mSystems.* 2019; **4**(4): e00280–19.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Link DHF: **Beschreibung Der Naturalien-Sammlung Der Universität Zu Rostock.** Adlers Erben, 1807.

[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](#)

Morel-Letellier I, Yuen B, Kück AC, *et al.*: **Adaptations to nitrogen availability drive ecological divergence of chemosynthetic symbionts.** *PLoS Genet.* 2024; **20**(5): e1011295.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Osvatic JT, Wilkins LG, Leibrecht L, *et al.*: **Global biogeography of chemosynthetic symbionts reveals both localized and globally distributed symbiont groups.** *Proc Natl Acad Sci U S A.* 2021; **118**(29): e2104378118.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Osvatic JT, Yuen B, Kunert M, *et al.*: **Gene loss and symbiont switching during adaptation to the deep sea in a globally distributed symbiosis.** *ISME J.* 2023; **17**(3): 453–466.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Parks DH, Imelfort M, Skennerton CT, *et al.*: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome Res.* 2015; **25**(7): 1043–1055.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Petersen JM, Kemper A, Gruber-Vodicka H, *et al.*: **Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation.** *Nat Microbiol.* 2016; **2**(1): 16195.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat Commun.* 2020; **11**(1): 1432.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Reynolds LK, Berg P, Zieman JC: **Lucinid clam influence on the biogeochemistry of the seagrass *Thalassia testudinum* sediments.** *Estuaries Coast.* 2007; **30**: 482–490.

[Publisher 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](#)

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](#)

Seemann T: **Prokka: rapid prokaryotic genome annotation.** *Bioinformatics.* 2014; **30**(14): 2068–2069.

[PubMed Abstract](#) | [Publisher Full Text](#)

Sieber CMK, Probst AJ, Sharrar A, *et al.*: **Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy.** *Nat Microbiol.* 2018; **3**(7): 836–843.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sogin EM, Kleiner M, Borowski C, *et al.*: **Life in the dark: phylogenetic and physiological diversity of chemosynthetic symbioses.** *Annu Rev Microbiol.* 2021; **75**: 695–718.

[PubMed Abstract](#) | [Publisher Full Text](#)

Suárez-Mozo NY, Papiol V, Enriquez C, *et al.*: **Molluscs along a salinity gradient in a hypersaline coastal lagoon, southern Gulf of Mexico.** *J Mar Biol Assoc U K.* 2023; **103**: e19.

[Publisher Full Text](#)

Sun J, Zhang Y, Xu T, *et al.*: **Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes.** *Nat Ecol Evol.* 2017; **1**(5): 121.

[PubMed Abstract](#) | [Publisher Full Text](#)

Taylor JD, Glover EA: **Cryptic diversity of chemosymbiotic bivalves: a systematic revision of worldwide *Anodontia* (Mollusca: Bivalvia: Lucinidae).** *Syst Biodivers.* 2005; **3**(3): 281–338.

[Publisher Full Text](#)

Taylor JD, Glover EA: **Biology, evolution and generic review of the chemosymbiotic bivalve family Lucinidae.** The Ray Society, 2021.

[Reference Source](#)

Taylor JD, Glover EA, Smith L, *et al.*: **New molecular phylogeny of Lucinidae: increased taxon base with focus on tropical Western Atlantic species (Mollusca: Bivalvia).** *Zootaxa.* 2016; **4196**(3): zootaxa.4196.3.2.

[PubMed Abstract](#) | [Publisher Full Text](#)

van der Heide T, Govers LL, de Fouw J, *et al.*: **A three-stage symbiosis forms the foundation of seagrass ecosystems.** *Science.* 2012; **336**(6087): 1432–1434.

[PubMed Abstract](#) | [Publisher 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](#)

Wilkins L, Yuen B, Petersen J, *et al.*: **The chromosomal genome sequence of the mollusc, *Ctena decussata* (O.G.Costa, 1829) and its bacterial endosymbiont *Candidatus Thiodiazotropha* sp. CDECU1 (Chromatiales) [version 1; peer review: 3 approved, 1 approved with reservations].** *Wellcome Open Res.* 2025; **10**: 435.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zhou C, Brown M, Blaxter M, *et al.*: **Oatk: a de novo assembly tool for complex plant organelle genomes.** *Genome Biol.* 2025; **26**(1): 235.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 28 March 2026

<https://doi.org/10.21956/wellcomeopenres.28530.r151560>

© 2026 Wilding C. 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.



Craig Wilding 

University College Dublin, Dublin, Ireland

The authors present a chromosome-level nuclear genome, mitochondrial genome, and four metagenome-assembled genomes (MAGs) from the bivalve *Anodonta alba*.

The rationale for sequencing this species is clear, and the methods are clearly presented and thus I have only some very minor queries/suggestions:

Whilst there is explicit mention in the text that gill tissue was used for Hi-C, there is no similar mention in the text (although it is given in Table 1) of which tissues were used for DNA extraction for PacBio or for RNA extraction. It would be helpful to make this clear in the text.

Hi-C methods: which REs were in the RE mastermix?

The genome sequence report states "GenomeScope2.0 analysis estimated the haploid genome size at 2 171.11 Mb, with a heterozygosity of 1.99% and repeat content of 55.18% (Figure 2)." However, the legend of Figure 2 appears to show the estimated length to be 2,167,172,647 bp which is different from that reported in the text. This may be my naivety in reading this plot, but is the figure in the text correct?

The BUSCO score of complete genes (83.8%) seems low for such a complete genome and importantly it does not meet the EBP benchmark (Table 4) - this really does need commenting upon in the Genome Sequence Report.

Table 5 - make clear which of these 4 metagenome assemblies was the one that did not meet the criteria for a MAG? Please clarify in the text or Table.

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 genetics

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 28 March 2026

<https://doi.org/10.21956/wellcomeopenres.28530.r151558>

© 2026 Abdelsalam 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.



Mohamed Abdelsalam 

Cairo University, Giza, Egypt

This Data Note presents a high-quality chromosomal-level genome assembly of *Anodontia alba*, a chemosymbiotic lucinid bivalve of considerable ecological and evolutionary importance. The assembly is technically sound, meets Earth BioGenome Project (EBP) reference standards, and the associated metagenome bins, including three high-quality MAGs, are valuable for the chemosymbiosis research community. The manuscript is well-structured and mostly clear. However, several specific scientific, methodological, and typographical concerns require attention before the final accepted version.

1. Single-Specimen Genome

The entire genome and metagenome assembly is derived from a single individual (xbAnoAlba1) collected from a single location (Sammy Creek Landing, FL, USA). While this is standard practice for genome data notes, the manuscript makes no acknowledgment or a Limitations subsection explicitly addressing single-specimen sampling and the implications of the 72.36% primary k-mer completeness figure.

2. BUSCO Completeness Below Benchmark — Insufficient Discussion

The BUSCO score is C:83.8% (S:82.2%, D:1.5%), with 12.2% missing genes against the *mollusca_odb10* reference set. This falls below the EBP benchmark of S > 90%. The manuscript notes this value in Table 4 but provides no narrative explanation for the relatively high proportion of missing BUSCO genes. Is the 12.2% missing BUSCO fraction attributable to repeat-rich assembly regions, gene prediction limitations, or genuine gene loss/divergence in *Anodontia alba*

compared to the mollusca_odb10 reference? The authors should at minimum discuss which explanation is most likely, even if empirically unresolved.

3. Symbiont Identification and Metagenome Report — Key Scientific Gap

The most scientifically significant metagenome bin is *Candidatus* Thiodiazotropha sp. (a gammaproteobacterial thioautotrophic endosymbiont, 99.39% complete, circularised, with 8,128× mean coverage), which is the known intracellular gill symbiont of *A. alba*.

-Does the *Candidatus* Thiodiazotropha sp. MAG recovered from this Florida specimen encode nitrogen fixation (*nif*) genes? Given that the Background explicitly mentions this as a variable trait across *A. alba* populations, this is a key scientific question the metagenome data could directly address.

-The two *Actinomyces* bins (*Rhodococcus qingshengii* and *Mycobacterium* sp.) are unusual residents for a marine bivalve gill. Are these confirmed as genuine gill-associated microbiota, or are they potential laboratory contaminants or environmental bacteria from the sediment? Their ecological significance in the context of *A. alba* biology is not discussed at all.

4. Specimen Identification Based Solely on Gross Morphology

The single specimen was "*identified by Diana Chin (Northeastern University) based on gross morphology*".

Confirming specimen identity using a molecular barcode (COI or 16S) or cross-referencing the assembled mitochondrial genome sequence (OZ024839.1, 18.48 kb) against GenBank would greatly strengthen taxonomic confidence.

-Was the assembled mitochondrial genome compared against existing *Anodontia alba* or congeneric sequences in GenBank or BOLD? This would provide a direct and inexpensive molecular confirmation of species identity using the data already generated.

5. Typographical and Language Issues

- Table 2, "Span (Mb)": The value is listed as 1 862.85 using a space as a thousands separator, while elsewhere in the text the same number is written as 1,862.85 or 1 862.85 megabases inconsistently. A single formatting convention should be used throughout.
- Table 4, k-mer completeness benchmark: The benchmark is stated as "≥ 95%" but the primary assembly achieves only 72.36%. The authors state the combined value is 98.20% and use this to justify meeting the standard, yet the table row shows the primary value prominently. A footnote clarifying that the combined assembly metric is used for EBP benchmarking would prevent confusion.
- Background, sentence: "*Genomes of Luciniscia nassula (Anderson et al., 2025), Ctena decussata (Wilkins et al., 2025) and Indoaustriella scarlati (Guo et al., 2025) have recently been published*" — the word "recently" is appropriate but could be more precisely replaced with "in 2025" for temporal clarity.
- Methods, Nucleic acid extraction: The sentence "*The xbAnoAlba1 sample was weighed and triaged to determine the appropriate extraction protocol*" uses the term "triaged" in an unusual non-clinical context. "Assessed" or "evaluated" would be more precise.

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: genetics

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.
