Check for updates

DATA NOTE

The genome sequence of the amphioxus, *Branchiostoma lanceolatum* (Pallas, 1774)

[version 1; peer review: 6 approved]

Patrick Adkins[1], John Bishop[1], Joanna Harley[1], Peter W. H. Holland[2],
Marine Biological Association Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]The Marine Biological Association, Plymouth, England, UK
[2]University of Oxford, Oxford, England, UK

v1 **First published:** 24 Feb 2025, **10**:95
https://doi.org/10.12688/wellcomeopenres.23671.1
Latest published: 24 Feb 2025, **10**:95
https://doi.org/10.12688/wellcomeopenres.23671.1

Abstract

We present a genome assembly from a specimen of *Branchiostoma lanceolatum* (Amphioxus; Chordata; Leptocardii; Amphioxiformes; Branchiostomatidae). The assembly contains two haplotypes with total lengths of 468.40 megabases and 465.81 megabases, respectively. Most of haplotype 1 (99.34%) is scaffolded into 19 chromosomal pseudomolecules. Haplotype 2 is a scaffold level assembly. The mitochondrial genome has also been assembled and is 15.14 kilobases in length.

Keywords

Branchiostoma lanceolatum, amphioxus, genome sequence, chromosomal, Amphioxiformes

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ✓ ✓ ✓ ✓ ✓ ✓

	1	2	3	4	5	6
version 1 24 Feb 2025	✓ view	✓ view	✓ view	✓ view	✓ view	✓ view

1. **Luohao Xu**, Southwest University, Chongqing, China
2. **Guang Li**, Xiamen University, Xiamen, China
3. **Salvatore D'Aniello**, Stazione Zoologica Anton Dohrn, Napoli, Italy
4. **Jr-Kai Yu**, Academia Sinica, Taipei, Taiwan
5. **Simona Candiani**, Università di Genova, Genoa, Italy
6. **Kun Wang**, Northwestern Polytechnical University, Xi'an, China

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: Adkins P: Investigation, Resources; **Bishop J**: Investigation, Resources, Writing – Review & Editing; **Harley J**: Investigation, Resources; **Holland PWH**: Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, https://doi.org/10.35802/206194] and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Adkins P *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: Adkins P, Bishop J, Harley J *et al.* **The genome sequence of the amphioxus, *Branchiostoma lanceolatum* (Pallas, 1774) [version 1; peer review: 6 approved]** Wellcome Open Research 2025, **10**:95
https://doi.org/10.12688/wellcomeopenres.23671.1

First published: 24 Feb 2025, **10**:95 https://doi.org/10.12688/wellcomeopenres.23671.1

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Cephalochordata; Leptocardii; Amphioxiformes; Branchiostomatidae; *Branchiostoma*; *Branchiostoma lanceolatum* (Pallas, 1774) (NCBI:txid7740)

Background

Lancelets or amphioxus are benthic filter-feeding marine invertebrates belonging to the subphylum Cephalochordata in the Phylum Chordata. Active animals capable of rapid movement, amphioxus have a stiff notochord running the length of the body, a dorsal nerve cord, pharyngeal slits and lateral segmented muscle blocks. These characters probably date back to the earliest chordates leading E.S. Goodrich to note "*Amphioxus is doubtless in some respects a very specialized animal . . . yet it preserves many primitive characters*" (Gaskell *et al.*, 1910). Several amphioxus species have been studied intensively as morphologically simple chordates and as outgroups for understanding the evolutionary origins of the vertebrates (D'Aniello *et al.*, 2023; Holland & Holland, 2022). These comparisons extend to the molecular level, with cephalochordates often retaining single-copy genes homologous to the multigene families of vertebrates generated by genome duplications (Furlong & Holland, 2002; Holland *et al.*, 2008).

There are over 20 species of cephalochordate worldwide but only one species, *Branchiostoma lanceolatum*, is recorded from Britain and Ireland. *B. lanceolatum* can be locally abundant in sand and gravel sediments off the coast of Devon and Cornwall, with scattered records from shallow seabed habitats around England, Wales and Scotland as far north as Shetland (NBN Atlas Partnership, 2024). A large population living in coarse shell gravel at 20 to 60 metres depth off the Eddystone reef south of Plymouth has been studied for over a century (Allen, 1899; Smith, 1932). There are also numerous records of planktonic amphioxus larvae sampled in Continuous Plankton Recorder surveys in the North Sea and Irish Sea (NBN Atlas Partnership, 2024), but these are difficult to identify to species level and the source populations not known. *B. lanceolatum* has a wide distribution in Europe, from the North Sea to the Atlantic coast of France and Portugal, and along the Mediterranean coastline of Spain, France and Italy (GBIF Secretariat, 2023). There are also records of *B. lanceolatum* from the Red Sea and Arabian Sea, possibly spread through the Suez Canal, but species identification has not been verified genetically (Maghsoudlou, 2018). Differences in spawning period, developmental rate and adult body size between Atlantic and Mediterranean populations raised questions over whether these are the same species; analyses of mitochondrial DNA sequences strongly support conspecificity (Caccavale *et al.*, 2021). The genetic or environmental basis for these morphological and life-history differences are not known, and most genetic work has been conducted on Mediterranean *B. lanceolatum* rather than on Atlantic populations.

Two high-quality genome sequences have been published from individual *B. lanceolatum* specimens collected in the Mediterranean Sea close to Argeles-sur-Mer, France (Brasó-Vives *et al.*, 2022; Marlétaz *et al.*, 2018). Here we report the complete genome sequence of an individual female *B. lanceolatum* collected from coastal sediments off Devon, UK. The genome sequence reported here will facilitate research into population genetics and local adaptations of *B. lanceolatum* in northern European waters and enable comparative analyses between Atlantic and Mediterranean *B. lanceolatum*. The genome sequence adds to a growing genomic data set for cephalochordates, a group of animals key for understanding the origins of vertebrates.

Genome sequence report

The genome of an adult *Branchiostoma lanceolatum* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating a total of 35.02 Gb (gigabases) from 2.97 million reads, providing approximately 75-fold coverage. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data, which produced 103.82 Gb from 687.58 million reads. Specimen and sequencing details are provided in Table 1.

The two haplotypes were combined for curation. Manual assembly curation corrected 51 missing joins or mis-joins and 22 haplotypic duplications. This reduced the scaffold number by 21.79%.

The final haplotype 1 assembly has a total length of 468.40 Mb in 121 sequence scaffolds, with 211 gaps, and a scaffold N50 of 24.11 Mb (Table 2). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.34%) of the assembly sequence was assigned to 19 chromosomal-level scaffolds. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 3).

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

The estimated Quality Value (QV) of the final haplotype 1 assembly is 66.0 and of haplotype 2 it is 66.4. The *k*-mer completeness of haplotype 1 is 61.36%, of haplotype 2 is 61.30%, and of the combined haplotypes is 99.84%. The BUSCO v5.4.3 completeness of haplotype 1 is estimated as 97.8% (single = 97.0%, duplicated = 0.8%), using the metazoa_odb10 reference set (n = 954). For haplotype 2, the BUSCO v5.4.3 completeness is 98.2% (single = 97.3%, duplicated = 0.9%), using the metazoa_odb10 reference set (n = 954).

Methods

Sample acquisition and DNA barcoding

An adult specimen of *Branchiostoma lanceolatum* (specimen ID MBA-220429-004A, ToLID klBraLanc2) was collected from H1 (Mv Sepia), Bigbury Bay, Devon, United Kingdom (latitude 50.26, longitude –3.95) on 2022-04-29, using a naturalist dredge deployed from the RV Sepia. The specimen was collected and identified by Patrick Adkins, John Bishop and

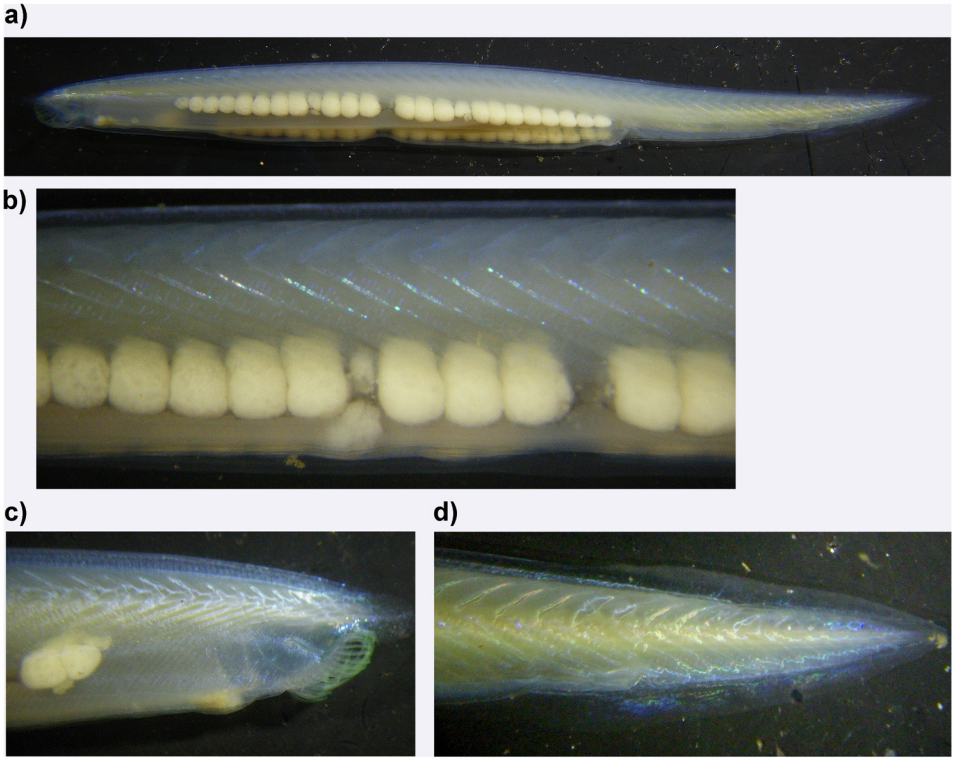

Figure 1. Photographs of the *Branchiostoma lanceolatum* (klBraLanc2) specimen used for genome sequencing. (**a**) Lateral view, anterior to the left. (**b**) Close-up of ripe gonads. (**c**) Close up of anterior showing oral cirri and pharyngeal slits. (**d**) Close up of tail showing chevron shaped lateral muscle blocks.

Table 1. Specimen and sequencing data for *Branchiostoma lanceolatum*.

Project information			
Study title	Branchiostoma lanceolatum (amphioxus)		
Umbrella BioProject	PRJEB74948		
Species	*Branchiostoma lanceolatum*		
BioSample	SAMEA111562176		
NCBI taxonomy ID	7740		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	klBraLanc2	SAMEA111562385	muscle
Hi-C sequencing	klBraLanc2	SAMEA111562383	muscle
RNA sequencing	klBraLanc2	SAMEA111562386	muscle
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Illumina NovaSeq 6000 (Hi-C)	ERR12945456	6.88e+08	103.82
Revio (PacBio)	ERR12921304	2.97e+06	35.02
Illumina NovaSeq X (RNA)	ERR13493941	8.88e+07	13.41

Table 2. Genome assembly data for the *Branchiostoma lanceolatum* assembly.

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	klBraLanc2.hap1.1	klBraLanc2.hap2.1
Assembly accession	GCA_964187965.1	GCA_964188055.1
Assembly level	chromosome	scaffold
Span (Mb)	468.4	465.8
Number of contigs	332	288
Number of scaffolds	121	100
Longest scaffold (Mb)	44.72	None
Assembly metrics (Benchmark)*	**Haplotype 1**	**Haplotype 2**
Contig N50 length (≥ 1 Mb)	3.23 Mb	3.63 Mb
Scaffold N50 length (= chromosome N50)	24.11 Mb	23.57 Mb
Consensus quality (QV) (≥ 40)	66.0	66.4
k-mer completeness	61.36%	61.30%
k-mer completeness combined (≥ 95%)	99.84%	
BUSCO** (S > 90%; D < 5%)	C:97.8%[S:97.0%,D:0.8%], F:1.2%,M:1.0%,n:954	C:98.2%[S:97.3%,D:0.9%], F:0.9%,M:0.9%,n:954
Percentage of assembly mapped to chromosomes (≥ 90%)	99.34%	-
Organelles (one complete allele)	Mitochondrial genome: 15.14 kb	-

* Assembly metric benchmarks are adapted from Rhie *et al.* (2021) and the Earth BioGenome Project Report on Assembly Standards September 2024.

** BUSCO scores based on the metazoa_odb10 BUSCO set using version 5.4.3. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

Joanna Harley (Marine Biological Association) and preserved on dry ice.

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 specimens and stored in ethanol, while the remaining parts were shipped on dry ice to the Wellcome Sanger Institute (WSI). The tissue was lysed, the COI marker region was amplified by PCR, and amplicons were sequenced and compared to the BOLD database, confirming the species identification (Crowley *et al.*, 2023). Following whole genome sequence generation, the relevant DNA barcode region was also used alongside the initial barcoding data for sample tracking at the WSI (Twyford *et al.*, 2024). The standard operating procedures for Darwin Tree of Life barcoding have been deposited on protocols.io (Beasley *et al.*, 2023).

Nucleic acid extraction

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory includes a sequence of core procedures: sample preparation and homogenisation, DNA extraction, fragmentation and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b). The klBraLanc2 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the muscle was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a).

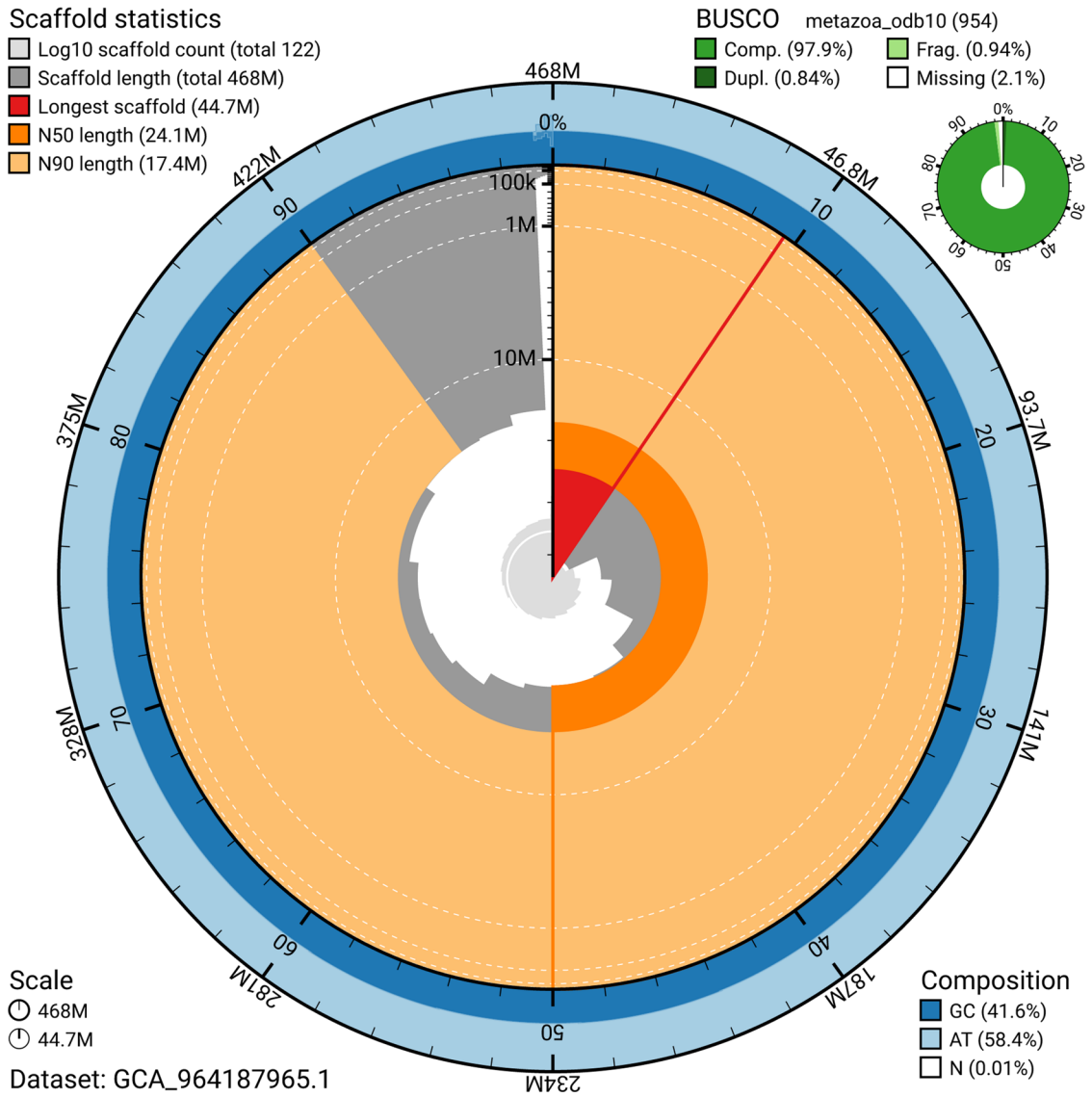

Figure 2. Genome assembly of *Branchiostoma lanceolatum*, klBraLanc2.hap1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around 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 metazoa_odb10 set is presented at the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Branchiostoma_lanceolatum/dataset/GCA_964187965.1/snail.

HMW DNA was extracted using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023a). DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Bates *et al.*, 2023). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Oatley *et al.*, 2023b). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from muscle tissue of klBraLanc2 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a

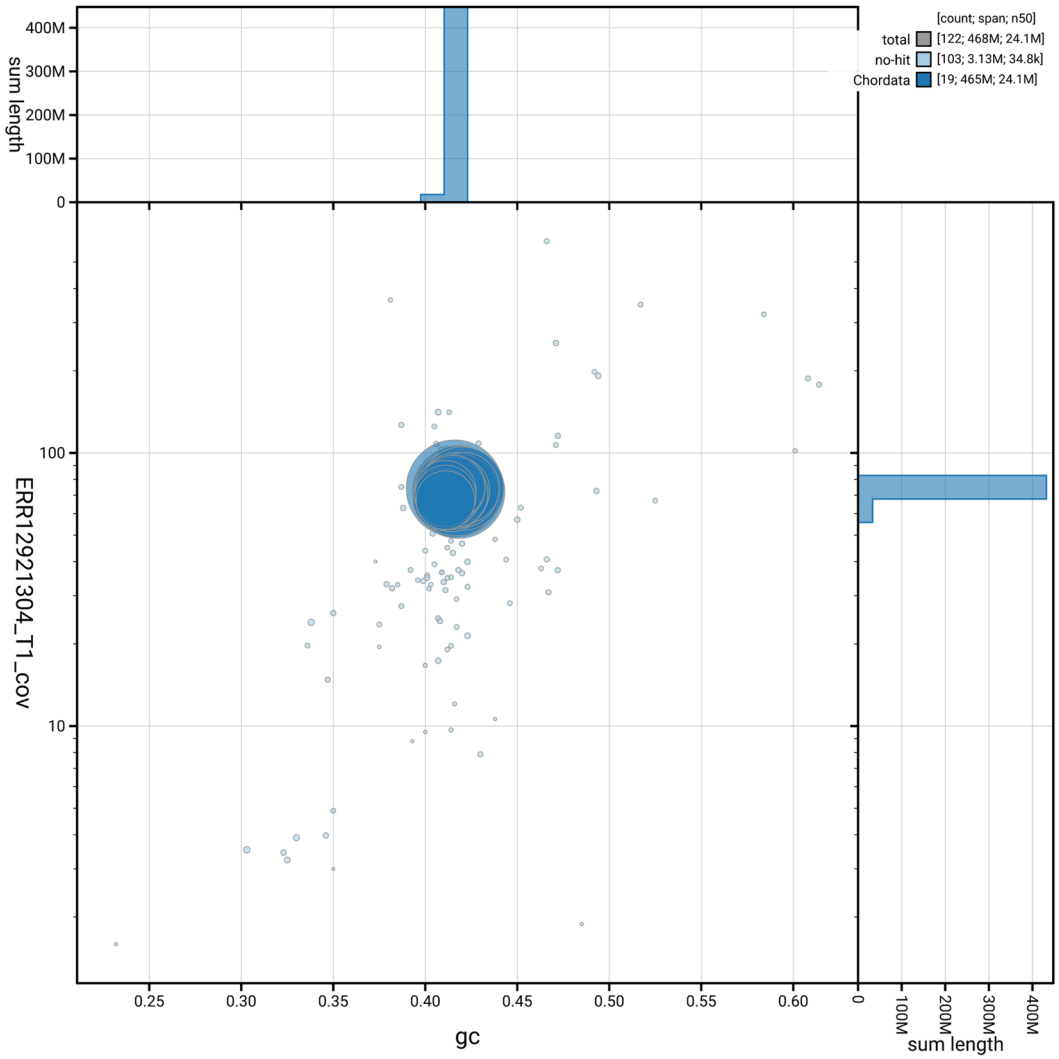

Figure 3. Genome assembly of *Branchiostoma lanceolatum*, klBraLanc2.hap1.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum 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 at https://blobtoolkit.genomehubs.org/view/Branchiostoma_lanceolatum/dataset/GCA_964187965.1/blob.

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.

Hi-C preparation

Muscle tissue from the klBraLanc2 sample was processed at the WSI Scientific Operations core, using the Arima-HiC v2 kit. Frozen tissue (stored at –80 °C) was fixed, and the DNA crosslinked using a TC buffer with 22% formaldehyde. After crosslinking, the tissue was homogenised using the Diagnocine Power Masher-II and BioMasher-II tubes and pestles. Following the kit manufacturer's instructions, crosslinked DNA was digested using a restriction enzyme master mix. The 5'-overhangs were then filled in and labelled with biotinylated nucleotides and proximally ligated. An overnight incubation was carried out for enzymes to digest remaining proteins and for crosslinks to reverse. A clean up was performed with SPRIselect beads prior to library preparation.

Library preparation and sequencing

PacBio HiFi

At a minimum, samples were required to have an average fragment size exceeding 8 kb and a total mass over 400 ng to proceed to the low input SMRTbell Prep Kit 3.0 protocol (Pacific Biosciences, California, USA), depending on genome size and sequencing depth required. Libraries were prepared using the SMRTbell Prep Kit 3.0 (Pacific Biosciences, California,

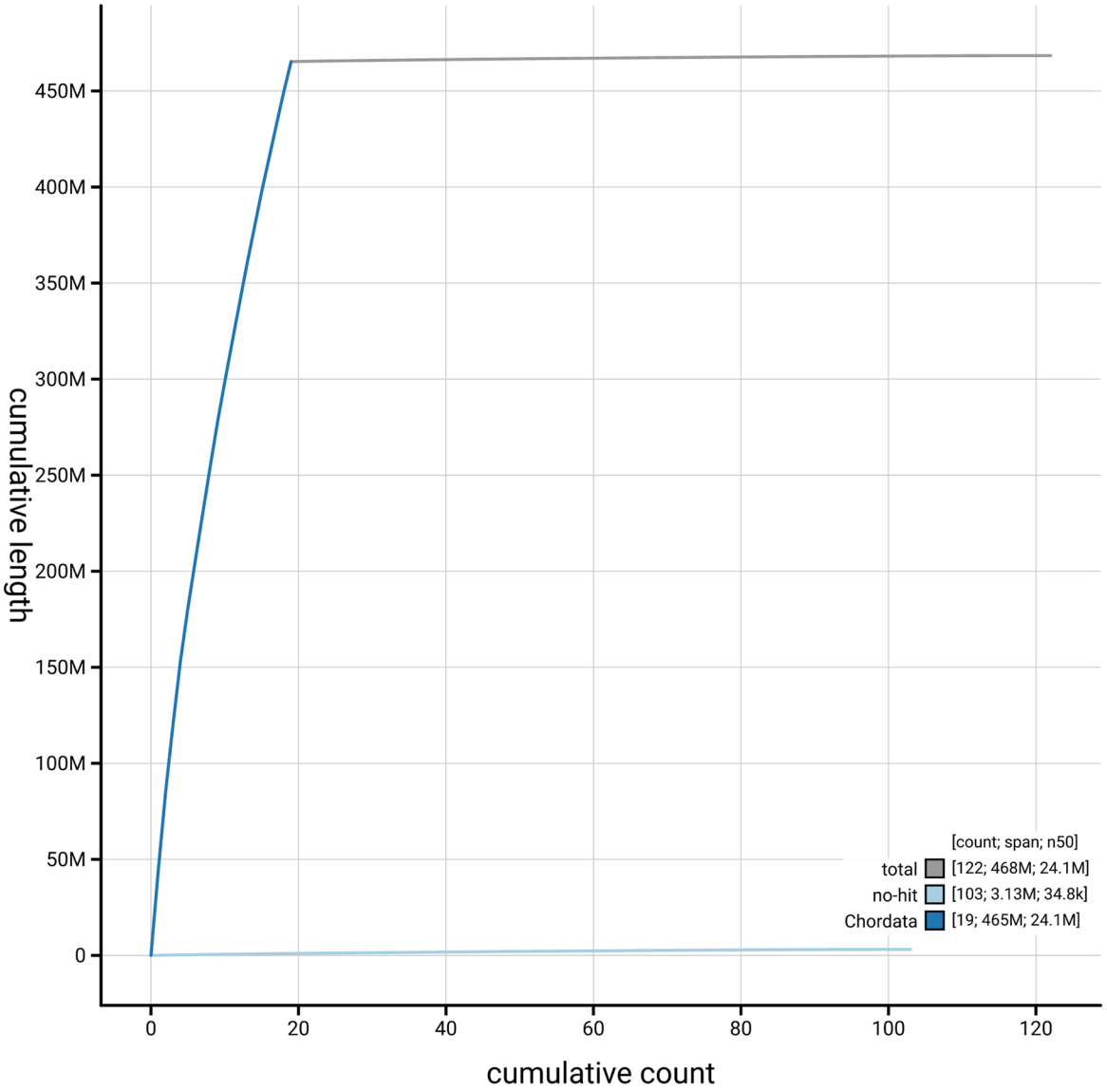

Figure 4. Genome assembly of *Branchiostoma lanceolatum* klBraLanc2.hap1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/Branchiostoma_lanceolatum/dataset/GCA_964187965.1/cumulative.

USA) as per the manufacturer's instructions. The kit includes the reagents required for end repair/A-tailing, adapter ligation, post-ligation SMRTbell bead cleanup, and nuclease treatment. Following the manufacturer's instructions, size selection and clean up was carried out using diluted AMPure PB beads (Pacific Biosciences, California, USA). DNA concentration was quantified using the Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with Qubit 1X dsDNA HS assay kit and the final library fragment size analysis was carried out using the Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) and gDNA 55kb BAC analysis kit.

Samples were sequenced on a Revio instrument (Pacific Biosciences, California, USA). Prepared libraries were normalised to 2 nM, and 15 µL was used for making complexes. Primers were annealed and polymerases were hybridised to create circularised complexes according to manufacturer's instructions. The complexes were purified with the 1.2X clean up with SMRTbell beads. The purified complexes were then diluted to the Revio loading concentration (in the range 200–300 pM), and spiked with a Revio sequencing internal control. Samples were sequenced on Revio 25M SMRT cells (Pacific Biosciences, California, USA). The SMRT link software, a PacBio web-based

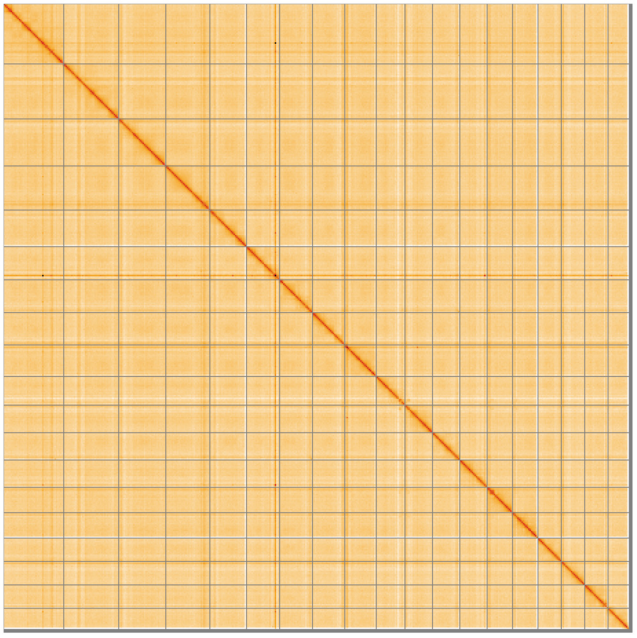

Figure 5. Genome assembly of *Branchiostoma lanceolatum* klBraLanc2.hap1.1: Hi-C contact map of the klBraLanc2.hap1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=TDyQj19wQX28exeWeDlF_Q.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Branchiostoma lanceolatum*, klBraLanc2, haplotype 1.

INSDC accession	Name	Length (Mb)	GC%
OZ076729.1	1	44.72	41.5
OZ076730.1	2	40.94	42
OZ076731.1	3	34.89	41.5
OZ076732.1	4	32.86	42
OZ076733.1	5	27.24	41.5
OZ076734.1	6	24.62	41.5
OZ076735.1	7	24.36	42
OZ076736.1	8	24.11	41.5
OZ076737.1	9	23.35	42
OZ076738.1	10	21.35	41.5
OZ076739.1	11	20.6	41.5
OZ076740.1	12	20.27	41.5
OZ076741.1	13	20.22	41
OZ076742.1	14	18.91	41.5
OZ076743.1	15	18.91	41.5
OZ076744.1	16	17.47	41
OZ076745.1	17	17.43	41
OZ076746.1	18	17.21	41
OZ076747.1	19	15.82	41
OZ076748.1	MT	0.02	38

end-to-end workflow manager, was used to set-up and monitor the run, as well as perform primary and secondary analysis of the data upon completion.

Hi-C

For Hi-C library preparation, DNA was fragmented using the Covaris E220 sonicator (Covaris) and size selected using SPRISelect beads to 400 to 600 bp. The DNA was then enriched using the Arima-HiC v2 kit Enrichment beads. Using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) for end repair, a-tailing, and adapter ligation. This uses a custom protocol which resembles the standard NEBNext Ultra II DNA Library Prep protocol but where library preparation occurs while DNA is bound to the Enrichment beads. For library amplification, 10 to 16 PCR cycles were required, determined by the sample biotinylation percentage. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq 6000 instrument.

RNA

Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit, following the manufacturer's

instructions. RNA sequencing was performed on the Illumina NovaSeq X instrument.

Genome assembly, curation and evaluation

Assembly

The HiFi reads were first assembled using Hifiasm (Cheng *et al.*, 2021; Cheng *et al.*, 2022) in Hi-C phasing mode, resulting in a pair of haplotype-resolved assemblies. The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). The contigs were further scaffolded using the provided Hi-C data (Rao *et al.*, 2014) in YaHS (Zhou *et al.*, 2023) using 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 (article in preparation). Flat files and maps used in curation were generated in TreeVal (Pointon *et al.*, 2023). Manual curation was primarily conducted using PretextView (Harry, 2022), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023) and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were corrected, and duplicate sequences were tagged and removed. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation (article in preparation).

Assembly quality assessment

The Merqury.FK tool (Rhie *et al.*, 2020), run in a Singularity container (Kurtzer *et al.*, 2017), was used to evaluate *k*-mer completeness and assembly quality for the primary and alternate haplotypes using the *k*-mer databases (k = 31) that were computed prior to genome assembly. The analysis outputs included assembly QV scores and completeness statistics.

A Hi-C contact map was produced for the final version of the assembly. The Hi-C reads were aligned using bwa-mem2 (Vasimuddin *et al.*, 2019) and the alignment files were combined using SAMtools (Danecek *et al.*, 2021). The Hi-C alignments were converted into a contact map using BEDTools (Quinlan & Hall, 2010) and the Cooler tool suite (Abdennur & Mirny, 2020). The contact map is visualised in HiGlass (Kerpedjiev *et al.*, 2018).

The blobtoolkit pipeline is a Nextflow (Di Tommaso *et al.*, 2017) port of the previous Snakemake Blobtoolkit pipeline (Challis *et al.*, 2020). It aligns the PacBio reads in SAMtools and minimap2 (Li, 2018) and generates coverage tracks for regions of fixed size. In parallel, it queries the GoaT database (Challis *et al.*, 2023) to identify all matching BUSCO lineages to run BUSCO (Manni *et al.*, 2021). For the three domain-level BUSCO lineages, the pipelinealigns the BUSCO genes to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) with DIAMOND blastp (Buchfink *et al.*, 2021). The genome is also divided into chunks according to the density of the BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND blastx. Genome sequences without a hit are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The blobtools suite combines all these 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), relying on the Conda package manager, the Bioconda initiative (Grüning *et al.*, 2018), the Biocontainers infrastructure (da Veiga Leprevost *et al.*, 2017), as well as the Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017) containerisation solutions.

Table 4 contains a list of relevant software tool versions and sources.

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**

Table 4. Software tools: versions and sources.

Software tool	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
BlobToolKit	4.3.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.4.3 and 5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond

Software tool	Version	Source
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	427104ea91c78c3b8b8b49f1a7d6bbeaa869ba1c	https://github.com/thegenemyers/FASTK
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	44086069ee7d4d3f6f3f0012569789ec138f42b84 aa44357826c0b6753eb28de	https://github.com/higlass/higlass
Merqury.FK	d00d98157618f4e8d1a9190026b19b471055b22e	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
NCBI Datasets	15.12.0	https://github.com/ncbi/datasets
Nextflow	23.04.0-5857	https://github.com/nextflow-io/nextflow
PretextView	0.2	https://github.com/sanger-tol/PretextView
purge_dups	1.2.5	https://github.com/dfguan/purge_dups
samtools	1.16.1, 1.17, and 1.18	https://github.com/samtools/samtools
sanger-tol/ascc	-	https://github.com/sanger-tol/ascc
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.0.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

Practice', which can be found in full on the Darwin Tree of Life website here. 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: Branchiostoma lanceolatum (amphioxus). Accession number PRJEB74948; https://identifiers.org/ena.embl/PRJEB74948. The genome sequence is released openly for reuse. The *Branchiostoma lanceolatum* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. 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.

Author information

Members of the Marine Biological Association Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.8382513.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.12158331.

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

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.12165051.

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

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

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

References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

Allen EJ: **On the fauna and bottom-deposits near the thirty-fathom line from the Eddystone grounds to start point.** *J Mar Biol Assoc U K.* 1899; **5**(4): 365–542.
Publisher Full Text

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Altschul SF, Gish W, Miller W, *et al.*: **Basic local alignment search tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

Bateman A, Martin MJ, Orchard S, *et al.*: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
PubMed Abstract | Publisher Full Text | Free Full Text

Bates A, Clayton-Lucey I, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for LI PacBio.** *protocols.io.* 2023.
Publisher Full Text

Beasley J, Uhl R, Forrest LL, *et al.*: **DNA barcoding SOPs for the Darwin Tree of Life project.** *protocols.io.* 2023; [Accessed 25 June 2024].
Publisher Full Text

Brasó-Vives M, Marlétaz F, Echchiki A: **Parallel evolution of amphioxus and vertebrate small-scale gene duplications.** *Genome Biol.* 2022; **23**(1): 243.
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

Caccavale F, Osca D, D'Aniello S, *et al.*: **Molecular Taxonomy Confirms that the Northeastern Atlantic and Mediterranean Sea Harbor a Single Lancelet, *Branchiostoma lanceolatum* (Pallas, 1774) (Cephalochordata: Leptocardii: Branchiostomatidae).** *PLoS One.* 2021; **16**(5): e0251358.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Kumar S, Sotero-Caio C, *et al.*: **Genomes on a Tree (GoaT): a versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic Tree of Life [version 1; peer review: 2 approved].** *Wellcome Open Res.* 2023; **8**: 24.
PubMed Abstract | Publisher Full Text | Free Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Jarvis ED, Fedrigo O, *et al.*: **Haplotype-resolved assembly of diploid genomes without parental data.** *Nat Biotechnol.* 2022; **40**(9): 1332–1335.
PubMed Abstract | Publisher Full Text | Free Full Text

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

da Veiga Leprevost F, Grüning BA, Alves Aflitos S, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017; **33**(16): 2580–2582.
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

D'Aniello S, Bertrand S, Escriva H: **Amphioxus as a model to study the evolution of development in chordates.** *eLife.* 2023; **12**: e87028.
PubMed Abstract | Publisher Full Text | Free Full Text

Denton A, Oatley G, Cornwell C, *et al.*: **Sanger Tree of Life sample homogenisation: PowerMash.** *protocols.io.* 2023a.
Publisher Full Text

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023b.
Publisher Full Text

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | Publisher Full Text

Diesh C, Stevens GJ, Xie P, *et al.*: **JBrowse 2: a modular genome browser with views of synteny and structural variation.** *Genome Biol.* 2023; **24**(1): 74.
PubMed Abstract | Publisher Full Text | Free Full Text

do Amaral RJV, Denton A, Yatsenko H, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.
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

Furlong RF, Holland PWH: **Were vertebrates octoploid?** *Philos Trans R Soc Lond B Biol Sci.* 2002; **357**(1420): 531–544.
PubMed Abstract | Publisher Full Text | Free Full Text

Gaskell WH, Macbride EW, Starling EH, *et al.*: **Discussion on the origin of vertebrates.** *Proceedings of the Linnaean Society of London.* 1910; **122**: 9–50.

GBIF Secretariat: ***Branchiostoma lanceolatum* (Pallas, 1774)**. 2023.

Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat Methods.* 2018; **15**(7): 475–476.
PubMed Abstract | Publisher Full Text | Free Full Text

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps**. 2022.
Reference Source

Holland LZ, Albalat R, Azumi K, *et al.*: **The amphioxus genome illuminates vertebrate origins and cephalochordate biology.** *Genome Res.* 2008; **18**(7): 1100–11.
PubMed Abstract | Publisher Full Text | Free Full Text

Holland LZ, Holland ND: **The invertebrate chordate amphioxus gives clues to vertebrate origins.** *Curr Top Dev Biol.* Academic Press, 2022; **147**: 563–594.
PubMed Abstract | 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

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
Publisher Full Text

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

Kurtzer GM, Sochat V, Bauer MW: **Singularity: scientific containers for mobility of compute.** *PLoS One.* 2017; **12**(5): e0177459.
PubMed Abstract | Publisher Full Text | Free Full Text

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**(18): 3094–3100.
PubMed Abstract | Publisher Full Text | Free Full Text

Maghsoudlou A: **Iranian *Branchiostoma* Species (Cephalochordata, Branchiostomatidae) Inhabiting Chabahar Bay (Gulf of Oman), with Remarks on Habitat Preferences.** *Turk J Zool.* 2018; **42**(2): 179–186.
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

Marlétaz F, Firbas PN, Maeso I, *et al.*: **Amphioxus functional genomics and the origins of vertebrate gene regulation.** *Nature.* 2018; **564**(7734): 64–70.
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, [Accessed 2 April 2024].
Reference Source

NBN Atlas Partnership: ***Branchiostoma lanceolatum* (Pallas, 1774) Lancelet.** 2024.
Reference Source

Oatley G, Denton A, Howard C: **Sanger Tree of Life HMW DNA extraction: automated MagAttract v.2.** *protocols.io.* 2023a.
Publisher Full Text

Oatley G, Sampaio F, Howard C: **Sanger Tree of Life fragmented DNA clean up: automated SPRI.** *protocols.io.* 2023b.
Publisher Full Text

Pointon DL, Eagles W, Sims Y, *et al.*: **sanger-tol/treeval v1.0.0 – Ancient Atlantis.** 2023.
Publisher Full Text

Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics.* 2010; **26**(6): 841–842.
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

Smith JE: **The Shell Gravel Deposits, and the Infauna of the Eddystone Grounds.** *J Mar Biol Assoc U K.* 1932; **18**(1): 243–278.
Publisher Full Text

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

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

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

Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔ ✔ ✔ ✔

Version 1

Reviewer Report 04 April 2025

https://doi.org/10.21956/wellcomeopenres.26114.r119573

© **2025 Wang K.** 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.

✔ **Kun Wang**
School of Ecology and Environment, Northwestern Polytechnical University, Xi'an, China

The manuscript is exceptionally well-written and clearly organized, with a logical flow that effectively guides readers through the research. The high-quality genome assembly presented here represents a valuable contribution to the scientific community, particularly for researchers studying cephalochordate evolution and population genetics in northern European waters. The methodological approach is comprehensively described, with appropriate technical details that would enable reproducibility. The authors have done an excellent job contextualizing their findings within existing literature, particularly in highlighting the significance of comparing Atlantic and Mediterranean *B. lanceolatum* populations.

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Comparative genomics, vertebrate evolution

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 18 March 2025

https://doi.org/10.21956/wellcomeopenres.26114.r119576

© 2025 Candiani S. 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.

Simona Candiani
Dipartimento di Scienze della Terra, dell'Ambiente e della Vita, Università di Genova, Genoa, Italy

The paper by Adkins and colleagues presents a data note on the genome sequence of a female *Branchiostoma lanceolatum* amphioxus collected in Devon, UK. This represents the first genome sequencing effort for an Atlantic population of this species.

Additionally, the mitochondrial sequence and transcriptomic data from muscle tissue have also been obtained.

These new data will be valuable for future annotation efforts, potentially enabling the community to generate new amphioxus genomes with improved annotations compared to previous versions.

In this regard, it would be interesting if the authors included a few lines discussing the quality of this sequencing compared to previous efforts for *Branchiostoma lanceolatum (Marletz et al., 2018; Simakov et al., 2020)*.

Overall, the report notes are well-written, and the technical procedures are clearly described.

References

1. Marlétaz F, Firbas PN, Maeso I, Tena JJ, et al.: Amphioxus functional genomics and the origins of vertebrate gene regulation.*Nature*. 2018; **564** (7734): 64-70 PubMed Abstract | Publisher Full Text
2. Simakov O, Marlétaz F, Yue JX, O'Connell B, et al.: Deeply conserved synteny resolves early events in vertebrate evolution.*Nat Ecol Evol*. 2020; **4** (6): 820-830 PubMed Abstract | Publisher Full Text

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: Developmental neuroscience and comparative genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 17 March 2025

https://doi.org/10.21956/wellcomeopenres.26114.r119572

© 2025 Yu J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Jr-Kai Yu
Academia Sinica, Taipei, Taiwan

In this paper, the authors report a new complete genome sequence and RNA-seq dataset from one single female adult of the amphioxus *Branchiostoma lanceolatum* collected from coastal sediments off Devon, UK. This new genome assembly will complement well to the existing genome dataset of *B. lanceolatum* collected from the Mediterranean population, and will be very useful for future comparative genomic studies with other cephalochordate species or with other groups of animals. Sufficient details of the methods and materials are provided in this paper, and the work flow and protocols appear to follow the standard operating procedures of a genome sequencing project. All raw sequence data and the assembly have been deposited in appropriate public domain.

I only have the following minor comments:

1. Regarding the RNA-seq dataset, in this paper the RNA was extracted only from the muscle tissue of the same adult sample. I am not sure whether this is because the standard procedure of the Darwin Tree of Life project requires the RNA sample to be collected from the same type of tissue for DNA sequencing. Generally, the muscle tissue contains a less diverse transcriptome profile and may not be able to provide an overall view of the expressed genes in an organism. If possible, collecting more tissue types, or some mixed-tissue type samples (such as cutting the entire animal into four pieces for RNA extraction individually) for RNA-seq may provide a more comprehensive dataset for analyzing the transcrioptome of this sample.

2. Page 10, 2nd paragraph of the right column, 8th line of this paragraph, the word “pipelinealigns” should be “pipeline aligns”.

3. Page 12, the reference list for “Brasó-Vives M, Marlétaz F, Echchiki A: Parallel evolution of amphioxus and…”; “*et al.*” should be added after the name “Echchiki A”, because there are more than three authors of this paper.

In addition, I notice that in previous Reviewer Report by Luohao Xu (posted on 27 February 2025, https://doi.org/10.21956/wellcomeopenres.26114.r119574), it was stated that “A *B. lanceolatum*

genome has also been published by Simakov et al.(2020)." Actually, the Simakov et al. (2020) paper is on the chromosome-scale genome assembly of the Florida amphioxus *B. floridae*, not *B. lanceolatum*.

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 Developmental Biology, Comparative Genomics, Chordate Evolution

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 17 March 2025

https://doi.org/10.21956/wellcomeopenres.26114.r119571

© 2025 D'Aniello S. 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.

Salvatore D'Aniello
Stazione Zoologica Anton Dohrn, Napoli, Italy

The report by Adkins and collaborators reports the complete genome sequence and assembly of *Branchiostoma lanceolatum* (Pallas, 1774) collected from Devon, UK. The Tree of Life consortium methodologies adhere to best genome assembly practices, and in particular, the use of PacBio HiFi long-read sequencing and Hi-C chromosome conformation capture have enabled the authors to achieve a high-quality, chromosome-level assembly. This resource will be highly valuable for comparative genomic analysis between *Branchiostoma lanceolatum* Atlantic and Mediterranean populations, a comparison that was previously not possible due to the availability of high-quality genome resources only from specimens collected in the Mediterranean Sea (southern France). This additional amphioxus genome will also be useful in future studies aimed at determining whether slight morphological differences among European amphioxus populations, as well as variations in breeding periods and developmental rates, could have a genetic or environmental

basis.
The rationale is clear, the datasets are accessible, and I have no major comments or concerns.

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: Evolution and Developmental Biology

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 13 March 2025

https://doi.org/10.21956/wellcomeopenres.26114.r119577

© 2025 Li G. 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.

Guang Li
Xiamen University, Xiamen, China

Complementary to previous studies that presented the genome sequences of a male and a gender-unknown *B. lanceolatum* from the Mediterranean sea, this study report the genome sequence of a female *B. lanceolatum* collected from coastal sediments off Devon, UK. It will be a valuable resource for research aiming to investigate population genetics and local adaptations of this lancelet species as the authors stated. I have several minor comments on it as following:

1) It would be useful to provide the quality information of the two haplotypes after initial assembly (before manual curation).

2) It would be better to offer an equivalent information for the haplotype 2 assembly as that described in the 3rd paragraph for the haplotype 1 in the report section.

3) Some comparisons between the assembly reported here and the one presented by Brasó-Vives et al. (2022) should be conducted.

4) Please make sure panel b in Figure 1 was taken from the same individual as panel a.

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 and developmental biology

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 27 February 2025

https://doi.org/10.21956/wellcomeopenres.26114.r119574

© 2025 Xu L. 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.

Luohao Xu
MOE Key Laboratory of Freshwater Fish Reproduction and Development, School of Life Sciences, Southwest University, Chongqing, China

Please make sure that the taxon information is corrrect: *Branchiostoma lanceolatum* (Amphioxus; Chordata; Leptocardii; Amphioxiformes; Branchiostomatidae)

The background part failed to cover more recent study using single-cell technology on amphioxus, such as Dai Y *et al.*(2024), Wu B *et al.*(2024) and Ma P *et al.* (2022)

A B. lanceolatum genome has also been published by Simakov *et al.(*2020). Please also include comparisons with other *Branchiostoma* genomes: Huang Z *et al.(2023)*

References
1. Dai Y, Zhong Y, Pan R, Yuan L, et al.: Evolutionary origin of the chordate nervous system revealed by amphioxus developmental trajectories.*Nat Ecol Evol*. 2024; **8** (9): 1693-1710 PubMed

Abstract | Publisher Full Text
2. Wu B, Xu W, Wu K, Li Y, et al.: Single-cell analysis of the amphioxus hepatic caecum and vertebrate liver reveals genetic mechanisms of vertebrate liver evolution.*Nat Ecol Evol*. 2024; **8** (10): 1972-1990 PubMed Abstract | Publisher Full Text
3. Ma P, Liu X, Xu Z, Liu H, et al.: Joint profiling of gene expression and chromatin accessibility during amphioxus development at single-cell resolution.*Cell Rep*. 2022; **39** (12): 110979 PubMed Abstract | Publisher Full Text
4. Simakov O, Marlétaz F, Yue JX, O'Connell B, et al.: Deeply conserved synteny resolves early events in vertebrate evolution.*Nat Ecol Evol*. 2020; **4** (6): 820-830 PubMed Abstract | Publisher Full Text
5. Huang Z, Xu L, Cai C, Zhou Y, et al.: Three amphioxus reference genomes reveal gene and chromosome evolution of chordates.*Proc Natl Acad Sci U S A*. 2023; **120** (10): e2201504120 PubMed Abstract | Publisher Full Text

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: 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.