DATA NOTE

REVISED The chromosomal genome sequence of the photosymbiotic ascidian, *Tridemnum clinides* Kott, 1977 and its associated microbial metagenome sequences

[version 2; peer review: 2 approved]

Euichi Hirose[1], Jose Victor Lopez[2], Graeme Oatley[3], Elizabeth Sinclair[3], Eerik Aunin[3], Noah Gettle[3], Camilla Santos[3], Michael Paulini[3], Haoyu Niu[3], Victoria McKenna[3], Rebecca O'Brien[3],
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

[1]Faculty of Science, University of the Ryukyus, Nishihara, Okinawa, Japan
[2]National Coral Reef Institute, Department of Biological Sciences, Nova Southeastern University, Fort Lauderdale, Florida, USA
[3]Tree of Life, Wellcome Sanger Institute, Hinxton, England, UK

v2 **First published:** 23 Jul 2025, **10**:357
https://doi.org/10.12688/wellcomeopenres.24563.1
Latest published: 06 Oct 2025, **10**:357
https://doi.org/10.12688/wellcomeopenres.24563.2

Abstract

We present a genome assembly from a specimen of *Tridemnum clinides* (photosymbiotic ascidian; Chordata; Ascidiacea; Aplousobranchia; Didemnidae). The *T. clinides* genome sequence has a total length of 906.92 megabases. Most of the assembly (97.83%) is scaffolded into 23 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 14.98 kilobases in length. The host ascidian has multiple symbionts, including *Prochloron*, a cyanobacterial genus that can also synthesise bioactive natural products of interest for potential therapeutic development. Biosynthesis of active compounds sometimes involves microbial associates.

Keywords

Tridemnum clinides, cyanobacterial symbionts, genome sequence, photosymbiosis, chromosomal, Aplousobranchia

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 06 Oct 2025		
version 1 23 Jul 2025	✓ view	✓ view

1. **Xavier Turon**, Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Blanes,, Spain

2. **Kevin M. Kocot**, University of Alabama,, Tuscaloosa,, USA

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: Hirose E: Writing – Original Draft Preparation, Writing – Review & Editing; **Lopez JV**: Resources, Supervision, Writing – Review & Editing; **Oatley G**: Investigation; **Sinclair E**: Investigation; **Aunin E**: Formal Analysis; **Gettle N**: Formal Analysis, Methodology; **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).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Hirose E, Lopez JV, Oatley G *et al.* **The chromosomal genome sequence of the photosymbiotic ascidian, *Trididemnum clinides* Kott, 1977 and its associated microbial metagenome sequences [version 2; peer review: 2 approved]** Wellcome Open Research 2025, **10**:357 https://doi.org/10.12688/wellcomeopenres.24563.2

First published: 23 Jul 2025, **10**:357 https://doi.org/10.12688/wellcomeopenres.24563.1

REVISED **Amendments from Version 1**

Version 2 of this article has the following changes in response to the peer reviewers' comments:

- We added text to the "Metagenome report" to link our results to previous observations.
- We added to the caption of Figure 1 to specify that it the photograph was taken in the laboratory.
- We added the storage temperature of the tissues in the "Sample acquisition" section and also specified in Table 1 that the whole colony was used for sequencing.

Any further responses from the reviewers can be found at the end of the article

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Tunicata; Ascidiacea; Aplousobranchia; Didemnidae; *Tridemnum*; *Tridemnum clinides* Kott, 1977 (NCBI:txid641091)

Background

In tropical and subtropical coral reef ecosystems, some colonial ascidians have evolved obligate symbiotic relationships with cyanobacteria, involving vertical transmission (Hirose, 2015). This ascidian–cyanobacterial symbiosis is the only known life-long photosymbiosis among chordate taxa. These symbiotic systems have attracted the attention of many pharmaceutical scientists for their secondary metabolites (Schmidt, 2015). The host ascidians are classified into four genera of the family Didemnidae (Tunicata: Ascidiacea: Aplousobranchia): *Didemnum*, *Tridemnum*, *Lissoclinum*, and *Diplosoma*. Among these, *Tridemnum* Della Valle, 1881 is distinguished by three rows of stigmata in the branchial sac and currently comprises 75 species (Shenkar *et al.*, 2025). However, molecular phylogenetic analysis based on COI gene sequences has not yet confirmed the monophyly of this genus (da Silva Oliveira *et al.*, 2017). At present, approximately ten *Tridemnum* species have been identified as photosymbiotic species.

Tridemnum clinides Kott, 1977 is a colonial ascidian recorded in Australia, French Polynesia, Eniwetak, Fiji, Guam, Philippines, Taiwan, Ryukyus (Okinawa), and Bonin Islands. Colonies harbor two types of unicellular cyanobacteria and one type of filamentous multicellular cyanobacteria in the cellulosic matrix of the host colony (i.e., tunic), whereas most photosymbiotic ascidian species harbour only one cyanobacterial species. The unicellular photosymbionts are regarded as *Prochloron* Lewin, 1977 and *Synechocystis* Sauvageau, 1892. However, the taxonomic affiliation of the filamentous cyanobacterium is uncertain. Few, if any, *Prochloron* species have ever been continuously grown in laboratory cultures, despite intense interest from various angles – e.g. microbial, biochemical (Lewin & Cheng, 1989; Rumengan *et al.*, 2020). Differences in the composition of photosynthetic pigments among the three suggested that each species is distributed differently within the host tunic, depending on the light microenvironment (Hirose *et al.*, 2009). The cushion-like colony is usually greyish in colour because of the dense distribution of calcareous spicules in the tunic. The spicules are less dense in shaded habitats, and the colony is often brownish because of cyanobacteria. While most symbiont cells are free from host cells and directly embedded in the matrix, some are engulfed by a tunic cell (haemocyte-like cells distributed in the tunic) and transferred to the host embryo brooded in the tunic (Kojima & Hirose, 2012). In this vertical transmission process, host cells are unlikely to discriminate between the three symbiont species, and single tunic cell often carries multiple species of cyanobacteria.

Complex material exchanges may occur among the host and the three symbiotic cyanobacteria, and multiple cyanobacterial species may allow ascidian colonies to retain a variety of toxic compounds and use them for chemical defence. Various ascidian and tunicate species have been known to hold or generate bioactive metabolites. For example, antiviral activities were identified in a few Caribbean *Tridemnum* species (Ramesh *et al.*, 2021)

Studying the whole genome assembly of this species and its cyanobacterial symbionts may provide an understanding of the molecular mechanisms and evolution of this multi-photosymbiotic system, including interspecific interactions among the cyanobacterial symbionts in the ascidian host.

Genome sequence report

Sequencing data

The genome of a specimen of *Tridemnum clinides* (Figure 1) was sequenced using Pacific Biosciences single-molecule HiFi long reads, generating 30.91 Gb from 3.56 million reads. Based on the estimated genome size, the sequencing data provided approximately 27× coverage of the genome. Chromosome conformation Hi-C data produced 87.23 Gb from 577.69 million reads. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing information.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC databases. The assembly was improved by manual curation, which corrected 325 misjoins or missing joins and removed 123 haplotypic duplications. These interventions reduced the total assembly length by 5.12%, decreased the scaffold count by 20.33%, and also decreased the scaffold N50 by 6.45%. The final assembly has a total length of 906.92 Mb in 191 scaffolds, with 646 gaps, and a scaffold N50 of 38.01 Mb (Table 2).

The snail plot in Figure 2 provides a summary of the assembly statistics, indicating the distribution of scaffold lengths and other assembly metrics. Figure 3 shows the distribution of scaffolds by GC proportion and coverage. Figure 4 presents a cumulative assembly plot, with separate curves representing different scaffold subsets assigned to various phyla, illustrating the completeness of the assembly.

Most of the assembly sequence (97.84%) was assigned to 23 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 5; Table 3).

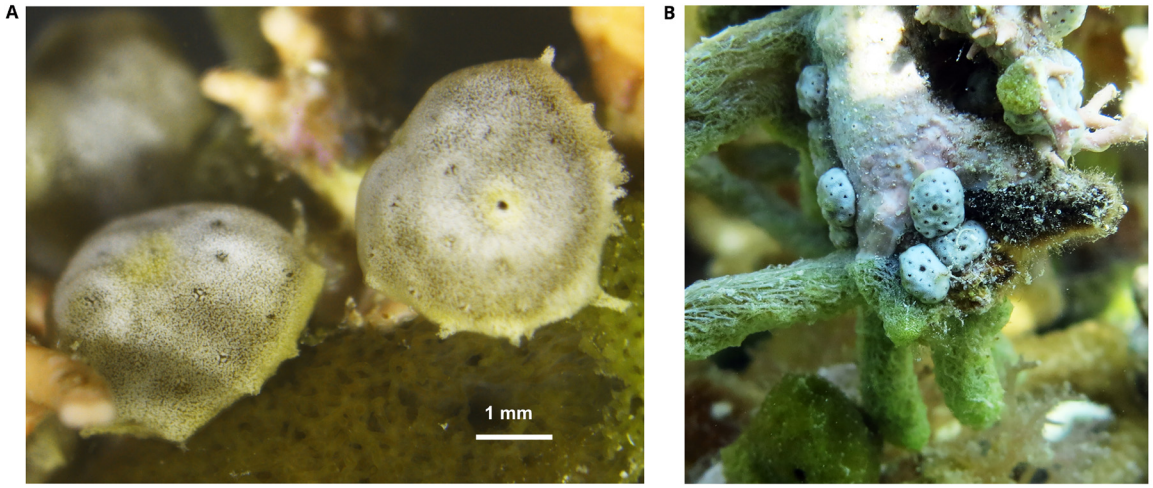

Figure 1. Photograph of the *Tridemnum clinides* (kaTriClin1) specimen used for genome sequencing. A. The specimen used for genome sequencing, photographed in the laboratory. **B**. *In situ* photograph of conspecific colonies neighbouring the sample colony.

Table 1. Specimen and sequencing data for *Tridemnum clinides*.

Project information			
Study title	Tridemnum clinides		
Umbrella BioProject	PRJEB66758		
Species	*Tridemnum clinides*		
BioSpecimen	SAMEA9873882		
NCBI taxonomy ID	641091		
Specimen information			
Technology	**ToLID**	**BioSample accession**	**Organism part**
PacBio long read sequencing	kaTriClin1	SAMEA9873897	Whole colony
Hi-C sequencing	kaTriClin1	SAMEA9873893	Whole colony
RNA sequencing	kaTriClin1	SAMEA9873895	Whole colony
Sequencing information			
Platform	**Run accession**	**Read count**	**Base count (Gb)**
Hi-C Illumina NovaSeq 6000	ERR12102423	5.78e+08	87.23
PacBio Sequel IIe	ERR12102457	5.97e+04	0.52
PacBio Sequel IIe	ERR12257376	2.24e+06	20.16
PacBio Sequel IIe	ERR12120136	1.26e+06	10.23
RNA Illumina NovaSeq X	ERR13093644	1.01e+07	1.53

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

Assembly quality metrics

The estimated Quality Value (QV) and *k*-mer completeness metrics, along with BUSCO completeness scores, were

Table 2. Genome assembly data for *Trididemnum clinides*.

Genome assembly		
Assembly name	kaTriClin1.1	
Assembly accession	GCA_963675345.1	
Alternate haplotype accession	*GCA_963675475.1*	
Assembly level for primary assembly	chromosome	
Span (Mb)	906.92	
Number of contigs	837	
Number of scaffolds	191	
Longest scaffold (Mb)	62.08	
Assembly metric	**Measure**	***Benchmark***
Contig N50 length	2.29 Mb	*≥ 1 Mb*
Scaffold N50 length	38.01 Mb	*= chromosome N50*
Consensus quality (QV)	Primary: 60.3; alternate: 60.5; combined 60.4	*≥ 40*
k-mer completeness	Primary: 74.13%; alternate: 72.83%; combined: 93.58%	*≥ 95%*
BUSCO*	C:91.9%[S:88.2%,D:3.7%],F:2.9%,M:5.1%,n:954	*S > 90%, D < 5%*
Percentage of assembly assigned to chromosomes	97.84%	*≥ 90%*
Organelles	Mitochondrial genome: 14.98 kb	*complete single alleles*

* BUSCO scores based on the metazoa_odb10 BUSCO set using version 5.5.0. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison.

calculated for each haplotype and the combined assembly. The QV reflects the base-level accuracy of the assembly, while *k*-mer completeness indicates the proportion of expected *k*-mers identified in the assembly. BUSCO scores provide a measure of completeness based on benchmarking universal single-copy orthologues.

The combined primary and alternate assemblies achieve an estimated QV of 60.4. The *k*-mer completeness is 74.13% for the primary haplotype and 72.83% for the alternate haplotype; and 93.58% for the combined primary and alternate assemblies. BUSCO v.5.5.0 analysis using the metazoa_odb10 reference set (n = 954) identified 91.9% of the expected gene set (single =88.2%, duplicated =3.7%).

Metagenome report

Sixteen binned genomes were generated from the metagenome assembly (Figure 6), of which 13 were classified as high-quality metagenome assembled genomes (MAGs) (see methods). The completeness values for these assemblies range from approximately 76% to 100% with contamination below 5%. A cladogram of the binned metagenomes is shown in Figure 7. The binned metagenomes include three cyanobacteria. Based on microscopic observations of the cyanobacterial composition (Hirose *et al.*, 2009), Prochloron sp. and Homoscilla sp., as shown in Figure 7 and Table 4, may correspond to Synechocystis sp. and multicellular filamentous cyanobacteria, respectively. For details on binned genomes see Table 4.

Methods

Sample acquisition

A *Trididemnum clinides* colony (specimen ID NSU00144301, ToLID kaTriClin1) was collected by snorkelling in the coral reef lagoon (<0.5 m in depth at low tide) of Bise, Okinawajima Island (Ryukyu Archipelago, Japan; latitude 26.71, longitude 127.88) on 2021-03-15. Colonies were attached to the basal portion of seagrass leaves (*Thalassia hemprichii*) or dead coral branches. The colonies were preserved in DMSO–NaCl solution (Dawson *et al.*, 1998). The samples were stored at 4°C until shipment and at ambient temperature during transport. They were then used for DNA extraction. Taxonomic identification was performed following Kott (2001). Neighbouring conspecific colonies were donated to Ryukyu University Museum Fujukan (RUMF-ZU-00069).

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 procedures: sample preparation and homogenisation, DNA extraction, fragmentation

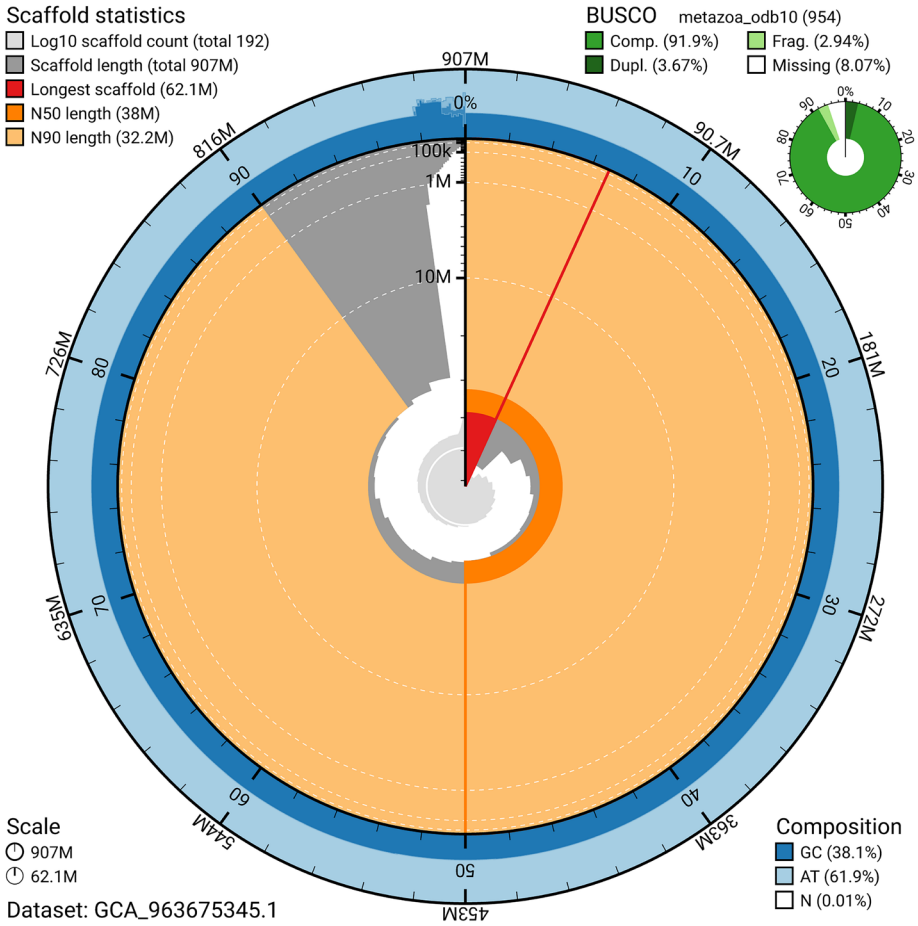

Figure 2. Genome assembly of *Tridemnum clinides*, kaTriClin1.1: metrics. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1,000 bins around the circumference. 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/GCA_963675345.1/dataset/GCA_963675345.1/snail.

and purification. Detailed protocols are available on protocols.io (Denton *et al.*, 2023b).

The kaTriClin1 sample was prepared for DNA extraction by weighing and dissecting it on dry ice (Jay *et al.*, 2023). Tissue from the animal was homogenised using a PowerMasher II tissue disruptor (Denton *et al.*, 2023a). HMW DNA was extracted using the Automated MagAttract v2 protocol (Oatley *et al.*, 2023). 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 (Strickland *et al.*, 2023). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from tissue of kaTriClin1 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 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.

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers'

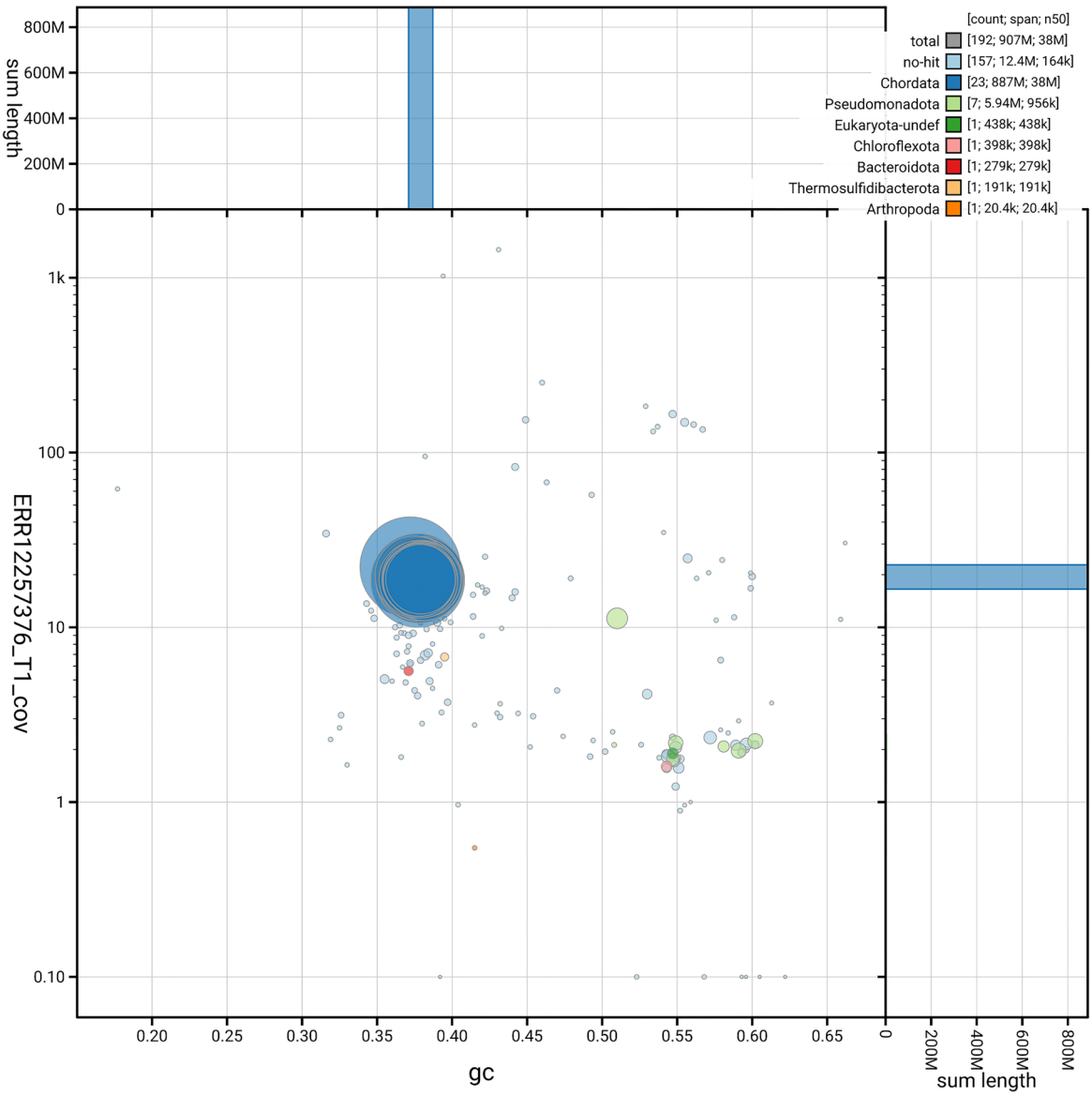

Figure 3. Genome assembly of *Trididemnum clinides*, kaTriClin1.1: BlobToolKit GC-coverage plot. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. 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/GCA_963675345.1/dataset/GCA_963675345.1/blob.

instructions. DNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences Sequel IIe. Tissue from the kaTriClin1 sample was processed for Hi-C sequencing at the WSI Scientific Operations core, using the Arima-HiC v2 kit and sequenced on the Illumina NovaSeq 6000 instrument. 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

Prior to assembly of the PacBio HiFi reads, a database of *k*-mer counts (k = 31) was generated from the filtered reads using

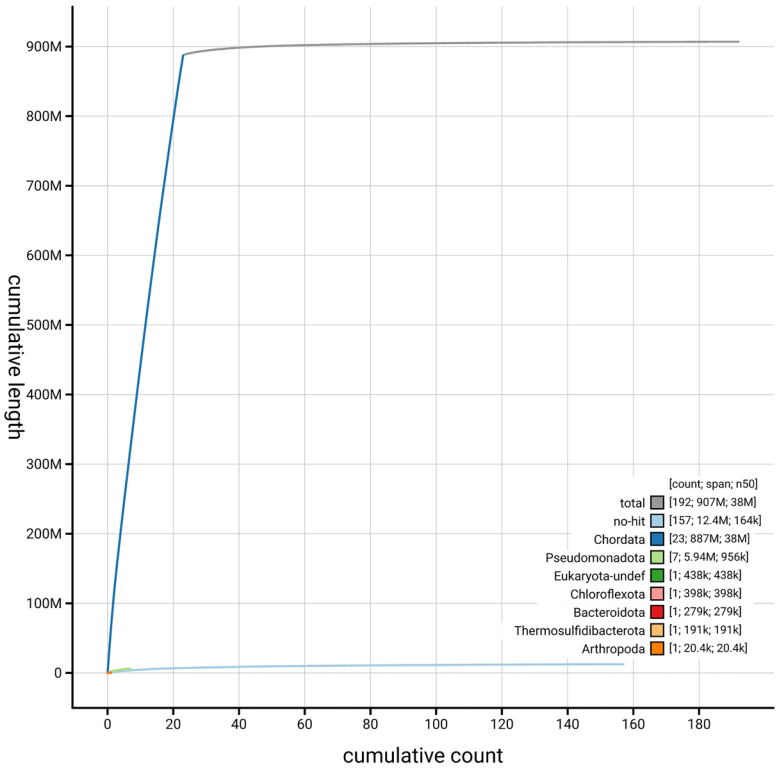

Figure 4. Genome assembly of *Trididemnum clinides*, kaTriClin1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/GCA_963675345.1/dataset/GCA_963675345.1/cumulative.

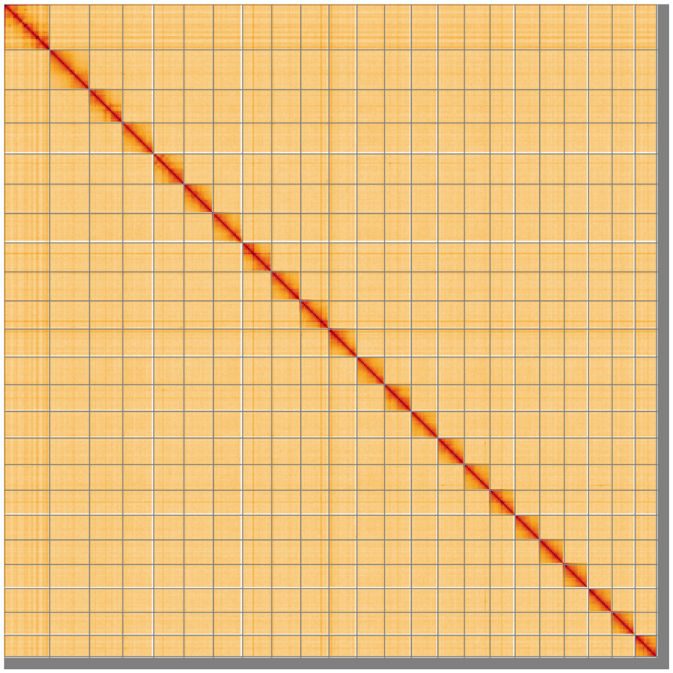

Figure 5. Genome assembly of *Trididemnum clinides*: Hi-C contact map of the kaTriClin1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=E4TYsVOJT8yDJ9jWfP0xFA.

Table 3. Chromosomal pseudomolecules in the genome assembly of *Tridemnum clinides*, kaTriClin1.

INSDC accession	Name	Length (Mb)	GC%
OY776284.1	1	62.08	37
OY776285.1	2	54.02	37.5
OY776286.1	3	45.32	38
OY776287.1	4	42.11	38
OY776288.1	5	41.04	38
OY776289.1	6	40.03	37.5
OY776290.1	7	39.83	38
OY776291.1	8	39.5	38
OY776292.1	9	39.46	38
OY776293.1	10	38.46	38
OY776294.1	11	38.01	37.5
OY776295.1	12	37.53	38
OY776296.1	13	36.46	38
OY776297.1	14	36.04	37.5
OY776298.1	15	36.0	38
OY776299.1	16	35.06	38
OY776300.1	17	33.93	38
OY776301.1	18	33.61	38
OY776302.1	19	33.31	38
OY776303.1	20	32.95	38
OY776304.1	21	32.23	38
OY776305.1	22	31.16	38
OY776306.1	23	29.14	38
OY776307.1	MT	0.01	17.5

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 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. Flat files and maps used in curation were generated via the TreeVal pipeline (Pointon *et al.*, 2023). Manual curation was conducted primarily in PretextView (Harry, 2022) and HiGlass (Kerpedjiev *et al.*, 2018), with additional insights provided by JBrowse2 (Diesh *et al.*, 2023). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Any identified contamination, missed joins, and mis-joins were amended, and duplicate sequences were tagged and removed. The curation process is documented at https://gitlab.com/wtsi-grit/rapid-curation.

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 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 pipeline aligns 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

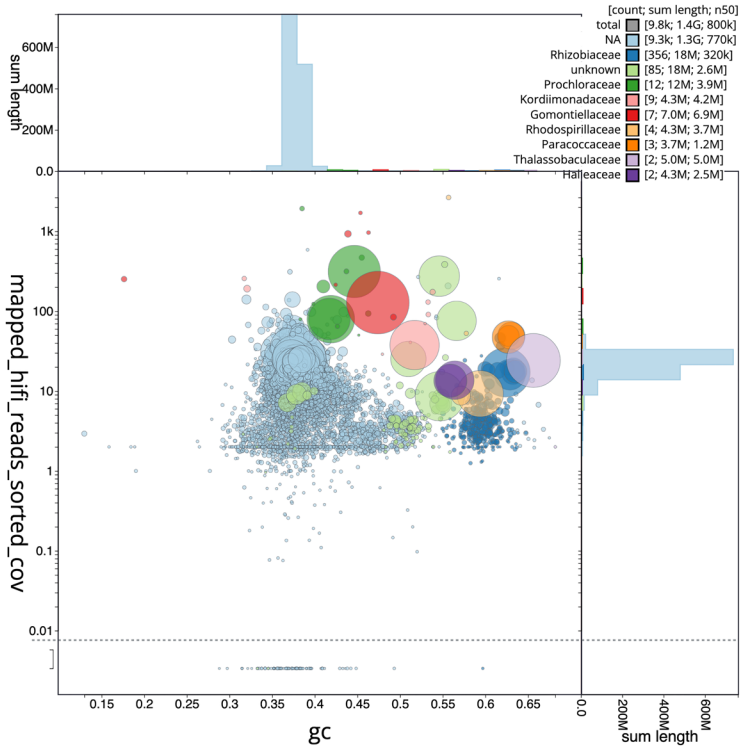

Figure 6. Blob plot of base coverage in mapped against GC proportion for sequences in the *Trididemnum clinides* metagenome. Binned contigs are coloured by family. Circles are sized in proportion to sequence length on a square-root scale, ranging from 510 to 6,879,066. Histograms show the distribution of sequence length sum along each axis.

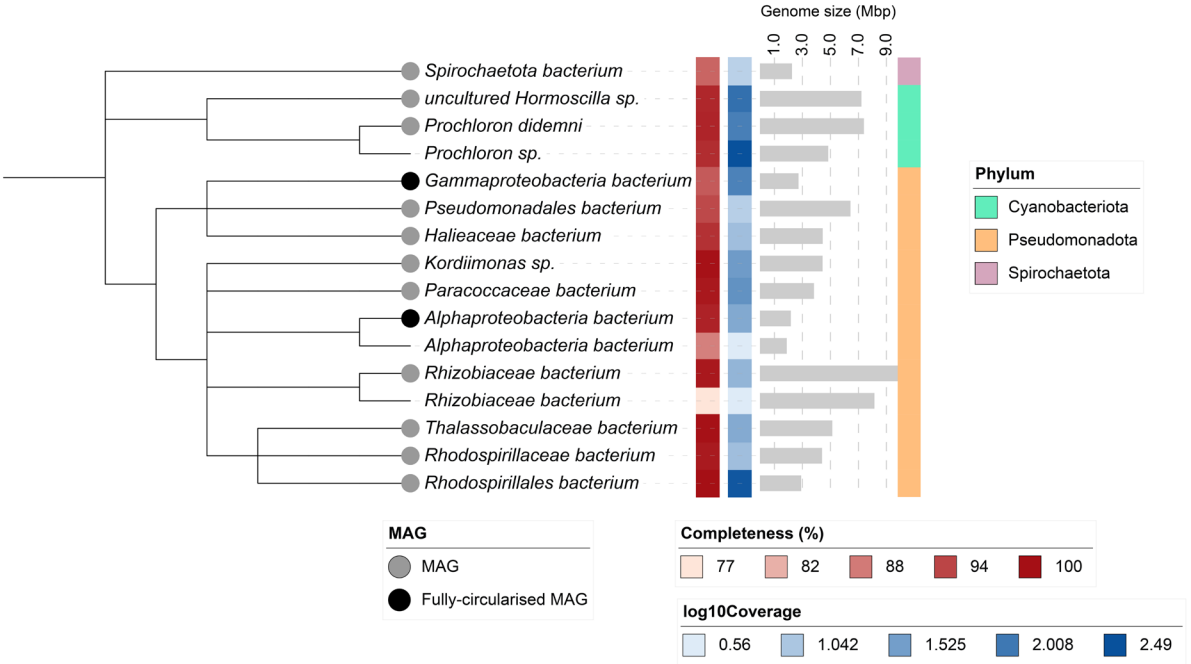

Figure 7. Cladogram showing the taxonomic placement of metagenome bins, constructed using NCBI taxonomic identifiers with taxonomizr and annotated in iTOL. Colours indicate phylum-level taxonomy. Additional tracks show sequencing coverage (log_{10}), genome size (Mbp), and completeness. Bins that meet the criteria for MAGs are marked with a grey circle; fully circularised MAGs are marked in black.

Table 4. Quality metrics and taxonomic assignments of the binned metagenomes.

NCBI taxon	Taxid	GTDB taxonomy	Quality	Size (bp)	Contigs	Circular	Mean coverage	Completeness (%)	Contamination (%)
Prochloron sp.	1215	g__Prochloron	Medium	4,812,471	1	Yes	311.97	96.81	1.78
Rhizobiaceae bacterium	1913961	g__SPNT01	Medium	8,113,228	345	No	3.65	76.96	3.22
Alphaproteobacteria bacterium	1913988	o__UBA11136	Medium	1,851,293	47	Partial	3.69	87.96	1.8
uncultured *Hormoscilla* sp.	1007660	g__Hormoscilla	High	7,175,611	7	Partial	127.67	97.53	1.71
Prochloron didemni	1216	s__Prochloron didemni	High	7,355,447	11	Partial	83.48	97.7	0.67
Rhizobiaceae bacterium	1913961	g__SPNT01	High	9,807,038	11	Partial	17.4	99.22	2.25
Paracoccaceae bacterium	1904441	g__JANSYL01	High	3,785,794	3	No	47.64	99.09	0.45
Rhodospirillales bacterium	2026786	f__JAJFIE01	High	2,884,309	2	Partial	269.29	100	1.49
Rhodospirillaceae bacterium	1898112	f__Rhodospirillaceae	High	4,366,830	4	Partial	13.12	99	0.75
Kordiimonas sp.	1970157	g__Kordiimonas	High	4,405,740	9	Partial	37.69	99.91	2.17
Thalassobaculaceae bacterium	3014067	f__Thalassobaculaceae	High	5,098,954	2	Partial	23.87	99.78	1.3
Alphaproteobacteria bacterium	1913988	o__UBA11136	High	2,141,931	1	Yes	25.11	98.05	0
Halieaceae bacterium	2026743	f__Halieaceae	High	4,424,697	2	No	13.56	96.45	1.34
Pseudomonadales bacterium	1891229	f__HTCC2089	High	6,398,238	15	No	8.24	94.03	4.29
Gammaproteobacteria bacterium	1913989	f__UBA4486	High	2,686,766	1	Yes	75.44	92.02	1.94
Spirochaetota bacterium	2202144	f__COTS27	High	2,220,260	19	No	7.99	90.95	1.12

Wellcome Open Research 2025, 10:357 Last updated: 06 OCT 2025

(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 5 contains a list of relevant software tool versions and sources.

Metagenome assembly

The metagenome assembly was generated using MetaMDBG (Benoit *et al.*, 2024) and binned using MetaBAT2 (Kang *et al.*, 2019), MaxBin (Wu *et al.*, 2014), bin3C (DeMaere & Darling, 2019) and MetaTOR. 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

Table 5. Software tools: versions and sources.

Software tool	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	4.3.3	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	1.2.1	https://github.com/Ecogenomics/CheckM
Cooler	0.8.11	https://github.com/open2c/cooler
DAS Tool	-	https://github.com/cmks/DAS_Tool
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
GoaT CLI	0.2.5	https://github.com/genomehubs/goat-cli
GTDB-TK	2.3.2	https://github.com/Ecogenomics/GTDBTk
Hifiasm	0.19.5-r587	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MaxBin	2.7	https://sourceforge.net/projects/maxbin/
MetaMDBG	1.1	https://github.com/GaetanBenoitDev/metaMDBG
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
MetaBat2	2.15-15-gd6ea400	https://bitbucket.org/berkeleylab/metabat/src/master/
MetaTOR	-	https://github.com/koszullab/metaTOR
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	3	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14, 1.17, and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.04.1	https://github.com/nextflow-io/nextflow
PretextView	0.2.5	https://github.com/sanger-tol/PretextView
PROKKA	1.14.5	https://github.com/vdejager/prokka
purge_dups	1.2.5	https://github.com/dfguan/purge_dups

Software tool	Version	Source
samtools	1.18	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.3.0	https://github.com/sanger-tol/blobtoolkit
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.2.0	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

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). The final bin set was filtered for bacteria and archaea. All bins were assessed for quality and categorised as metagenome-assembled genomes (MAGs) if they met the following criteria: contamination ≤ 5%, presence of 5S, 16S, and 23S rRNA genes, at least 18 unique tRNAs, and either ≥ 90% completeness or ≥ 50% completeness with fully circularised chromosomes. Bins that did not meet these thresholds, or were identified as taxonomic replicates of MAGs, were retained as 'binned metagenomes' provided they had ≥ 50% completeness and ≤ 10% contamination. A cladogram based on NCBI taxonomic assignments was generated using the 'taxonomizr' package in R. The tree was visualised and annotated using iTOL (Letunic & Bork, 2024). Software tool versions and sources are given in Table 5.

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 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: Trididemnum clinides. Accession number PRJEB66758; https://identifiers.org/ena.embl/PRJEB66758. The genome sequence is released openly for reuse. The *Trididemnum clinides* genome sequencing initiative is part of the Aquatic Symbiosis Genomics (ASG) project (https://www.ebi.ac.uk/ena/browser/view/PRJEB43743). 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 Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory Team are listed here: https://doi.org/10.5281/zenodo.13960213.

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

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.15495178.

Members of the European Bioinformatics Institute ASG Data Portal team are listed here: https://doi.org/10.5281/zenodo.10076466.

Members of the Wellcome Sanger Institute/Aquatic Symbiosis Genomics Project Leadership are listed here: https://doi.org/10.5281/zenodo.10184833.

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

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

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

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

da Silva Oliveira FA, Michonneau F, da Cruz Lotufo TM: **Molecular phylogeny of Didemnidae (Ascidiacea: Tunicata).** *Zool J Linn Soc.* 2017; **180**(3): 603–612.
Publisher 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

Dawson MN, Raskoff KA, Jacobs DK: **Field preservation of marine invertebrate tissue for DNA analyses.** *Mol Mar Biol Biotechnol.* 1998; **7**(2): 145–152.
PubMed Abstract

DeMaere MZ, Darling AE: **bin3C: exploiting Hi-C sequencing data to accurately resolve metagenome-assembled genomes.** *Genome Biol.* 2019; **20**(1): 46.
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

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

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

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

Hirose E: **Ascidian photosymbiosis: diversity of cyanobacterial transmission during embryogenesis.** *Genesis.* 2015; **53**(1): 121–131.
PubMed Abstract | Publisher Full Text

Hirose E, Uchida H, Murakami A: **Ultrastructural and microspectrophotometric characterization of multiple species of cyanobacterial photosymbionts coexisting in the colonial ascidian *Tridemnum clinides* (Tunicata, Ascidiacea, Didemnidae).** *Eur J Phycol.* 2009; **44**(3): 365–375.
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

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

Kojima A, Hirose E: **Transmission of cyanobacterial symbionts during embryogenesis in the coral reef ascidians *Tridemnum nubilum and T. clinides* (Didemnidae, Ascidiacea, Chordata).** *Biol Bull.* 2012; **222**(1): 63–73.
PubMed Abstract | Publisher Full Text

Kott P: **The Australian Ascidiacea, Part 4, Aplousobranchia (3), Didemnidae.** *Memoirs of the Queensland Museum.* 2001; **47**: 1–407.
Reference Source

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

Letunic I, Bork P: **Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool.** *Nucleic Acids Res.* 2024; **52**(W1): W78–W82.
PubMed Abstract | Publisher Full Text | Free Full Text

Lewin RA, Cheng L, (eds.): **Prochloron: a microbial enigma.** Boston, MA: Springer US, 1989.
Publisher Full Text

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

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | Publisher Full Text | Free Full Text

Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2. [Accessed 2 April 2024].
Reference Source

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

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): 1043–55.
PubMed Abstract | Publisher Full Text | Free 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**

Ramesh C, Tulasi BR, Raju M, *et al.*: **Marine natural products from tunicates and their associated microbes.** *Mar Drugs.* 2021; **19**(6): 308.
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**

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**

Rumengan IFM, Kubelaborbir TM, Tallei TE: **Data on the cultivation of *Prochloron* sp. at different salinity levels.** *Data Brief.* 2020; **29**: 105241.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Schmidt EW: **The secret to a successful relationship: lasting chemistry between ascidians and their symbiotic bacteria.** *Invertebr Biol.* 2015; **134**(1): 88–102.
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**

Shenkar N, Gittenberger A, Lambert G, *et al.*: **Ascidiacea world database**. 2025.
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**

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023.
Publisher 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

Wu YW, Tang YH, Tringe SG, *et al.*: **MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm.** *Microbiome.* 2014; **2**(1): 26.
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 12 September 2025

https://doi.org/10.21956/wellcomeopenres.27072.r128350

✔ **Kevin M. Kocot**
[1] Department of Biological Sciences,, University of Alabama,, Tuscaloosa,, AL,, USA
[2] Department of Biological Sciences,, University of Alabama,, Tuscaloosa,, AL,, USA

In this manuscript, Hirose et al. describe the chromosomal genome assembly of a photosymbiotic ascidian and the genomes of associated prokaryotes. I found all aspects of the manuscript to be thorough and appropriately detailed. The manuscript is generally very well-written. Overall, I think the manuscript is suitable for publication as-is but see a few minor comments below:

- Abstract: If space permits, it might be worthwhile to clarify that Prochloron is a cyanobacterial symbiont.
- In the Keywords and Data availability sections, Tridemnum clinides should be italicized.
- Background: There is a grammatical issue (I think a word is missing) in the first sentence of the last paragraph of this section.
- Sequencing data section of Genome sequence report section: There is an unnecessary hyphen in “con-formation” in the second-to-last paragraph of this section.
- Table 1: Could any more specific details be provided about what tissue(s)/material were used? Was any tunic material included? If the gut was intentionally excluded, that might be valuable to note.
- Methods: Please note the storage temperature of the tissues stored in DMSO-NaCl buffer and approximately how long it was stored prior to nucleic acid extraction. Please also clarify that the tissue used for RNA extraction was also preserved in this manner.
- Figure 7: Words like “uncultured” and “sp.” do not need to be italicized.

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

Are the protocols appropriate and is the work technically sound?
No

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

Yes

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

Competing Interests: No competing interests were disclosed.

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

Reviewer Report 30 August 2025

https://doi.org/10.21956/wellcomeopenres.27072.r128359

Xavier Turon
[1] Department of Marine Ecology, Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Blanes,, Spain
[2] Department of Marine Ecology, Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Blanes,, Spain

I have read this article with interest and, in my view, it provides new and relevant information. The format is the same I see in other Wellcome Open Research papers, as are the protocols followed, graphs, and tables. So, I imagine these are standard for this type of publication and I am not arguing with them.

The parameters of the assembly of the genome and associated metagenomes are correct.

I want to emphasize that T. clinides is a well-known model, examined from morphological and biological points of view previously by the first author.

What I miss (but I don't know if it falls in the scope of this type of reports) is a link between previous observations and the metagenome results. The species is known to harbor three main cyanobacterial species, likely a Prochloron, a Synechocystis, and a filamentous cyanobacterium. How do this relate to the 16 binned metagenomes? Do they confirm the identity of the vertically transmitted symbionts? I see no Synechocystis among the MAGs, is the Hormoscilla consistent with the filamentous unidentified symbiont?

I would also have appreciated more comments, at least on how the chromosomal number inferred related to that found in other didemnids, or about the mitochondrial genome (number of genes, comparison with closely related groups...). But, as said, I infer that this is the format of this type of notes, and that this information can be further elaborated in future works.

Picture 1A does not seem an in situ image, please
clarify in legend if it has been taken at the lab.

Typos:

“Few, if any Prochloron species have ever been continuously…” should be “Few, if any, Prochloron…”

Close quotation mark in: Differences in the composition of photosynthetic pigments among the three suggested that each “species is distributed differently”

Correct: Avail-able in public sequence repositories

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: Marine Biology, Ascidians

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.