



DATA NOTE

The genome sequence of a coronate scyphozoan jellyfish, *Nausithoe racemosa* (Komai, 1936) (Coronatae: Nausithoidae), and a metagenome-assembled genome of the associated cyanobacterium *Moorena producens*

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

Michael N Dawson ¹, Graeme Oatley ², Elizabeth Sinclair ², Eerik Aunin², Noah Gettle ², Camilla Santos², Michael Paulini ², Haoyu Niu², Victoria McKenna², Rebecca O'Brien ², 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

¹Life & Environmental Sciences, University of California, Merced, Merced, California, USA

²Tree of Life Programme, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 28 Jul 2026, 11:476
<https://doi.org/10.12688/wellcomeopenres.27239.1>
 Latest published: 28 Jul 2026, 11:476
<https://doi.org/10.12688/wellcomeopenres.27239.1>

Abstract

We present a genome assembly from a specimen of *Nausithoe racemosa* (coronate scyphozoan jellyfish; Cnidaria; Scyphozoa; Coronatae; Nausithoidae). The assembly contains two haplotypes with total lengths of 4 784.66 megabases and 4 868.20 megabases. Most of haplotype 1 (97.34%) is scaffolded into 20 chromosomal pseudomolecules. Haplotype 2 was assembled to scaffold level. The mitochondrial genome has also been assembled, with a length of 13.97 kilobases. From the metagenome data, we recovered one high-quality metagenome-assembled genome.

Keywords

Nausithoe racemosa, coronate scyphozoan jellyfish, genome sequence, chromosomal, Coronatae, microbial metagenome assembly

Open Peer Review

Approval Status ? ? ✓

	1	2	3
version 1	?	?	✓
28 Jul 2026	view	view	view
1. Bingpeng Xing , Third Institute of Oceanography Ministry of Natural Resources, Xiamen, China			
2. Nicholas S Kron , University of Miami, Miami, USA			
3. Shawn M Burgess , National Human Genome Research Institute, Bethesda, USA			



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

Any reports and responses or comments on the article can be found at the end of the article.

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

Author roles: **Dawson MN:** Conceptualization, Funding Acquisition, Investigation, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Oatley G:** Investigation; **Sinclair E:** Investigation; **Aunin E:** Formal Analysis; **Gettle N:** Formal Analysis; **Santos C:** Formal Analysis; **Paulini M:** Formal Analysis; **Niu H:** Project Administration; **McKenna V:** Project Administration; **O'Brien R:** Project Administration;

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by the Gordon and Betty Moore Foundation through a grant (GBMF8897) to the Wellcome Sanger Institute to support the Aquatic Symbiosis Genomics Project, and by Wellcome through core funding to the Wellcome Sanger Institute (220540). Collections for this project were funded by the US National Science Foundation grant "RAPID: Discovering Global Diversity in Pelagic Symbioses (Vessels of Opportunity)" to Michael N Dawson, DEB-2132455.

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

Copyright: © 2026 Dawson MN *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: Dawson MN, Oatley G, Sinclair E *et al.* **The genome sequence of a coronate scyphozoan jellyfish, *Nausithoe racemosa* (Komai, 1936) (Coronatae: Nausithoidae), and a metagenome-assembled genome of the associated cyanobacterium *Moorena producens* [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2026, 11:476 <https://doi.org/10.12688/wellcomeopenres.27239.1>

First published: 28 Jul 2026, 11:476 <https://doi.org/10.12688/wellcomeopenres.27239.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Cnidaria; Scyphozoa; Coronatae; Nausithoidae; *Nausithoe*; *Nausithoe racemosa* (Komai, 1936) (NCBI:txid2976324).

Background

The coronate medusae (Cnidaria: Scyphozoa: Coronatae) comprise one of two major clades of “true jellyfish” within Scyphozoa, the other being Discomedusae (Bayha *et al.*, 2010). Coronates are best known as deep-water, mesopelagic jellyfish (Arai, 1997: p. 5; Mianzan & Cornelius, 1999; OBIS, 2026). Shallow-water species are exceptions, notably in *Nausithoe* and *Linuche* (Bayha *et al.*, 2010; Molinari *et al.*, 2023). These genera are also unusual among coronates because both include photosymbiotic species (Djeghri *et al.*, 2019; Hingston *et al.*, 2007).

The phylogeny of coronates is poorly resolved, leaving several questions unanswered (Bayha *et al.*, 2010). For example: 1. Is Nausithoidae basally branching in the clade? 2. Is *Nausithoe* polyphyletic, with one lineage inhabiting deep-water oceanic habitats and the other inhabiting shallow-water coastal habitats? 3. Are the shallow-water coronates, *Nausithoe* and *Stephanoscyphus* monophyletic? 4. Did the photosymbioses in *Nausithoe* and *Stephanoscyphus* originate independently?

The origin of photosymbiosis in *Nausithoe*, whether or not it is independent of that in *Stephanoscyphus*, is noteworthy because it is clearly independent of another origin of photosymbiosis within the Scyphozoa: in the Rhizostomeae clade Kolpophorae (Djeghri *et al.*, 2019). In both *Nausithoe* and Kolpophorae, the photosymbiont is a member of the Symbiodiniaceae (Dinoflagellata), often referred to as zooxanthellae. In *Nausithoe*, preliminary sequencing analyses indicate that the zooxanthellar endosymbionts belong to the genus *Cladocopium* (previously *Symbiodinium*) (LaJeunesse *et al.*, 2018; Higgins-Poling, 2024: p. 32). The holobiont is too poorly studied to say whether *Nausithoe* may form photosymbioses with additional species of Symbiodiniaceae, or whether a single host can harbour more than one dinoflagellate species at a time. These patterns are known among scyphozoans (Trench & Luong-van Thinh, 1995; Arai, 1997: p. 105).

Nausithoe is also relevant to questions about life-history evolution in the clade. If Nausithoidae is relatively basal in the clade, close to Linuchidae, this would support the inference that a polyp stage is plesiomorphic for Coronatae, consistent with ancestral state reconstruction for most medusozoans, and that polyps ancestrally reproduce asexually through polydisc strobilation (Collins, 2002). In that case, the benthic polyp stage was lost from derived deep-water coronates such as *Periphylla* (Helm, 2018; Jarms *et al.*, 1999). Nausithoid polyps are dendritic in form, often bushy (Figure 1), and produce many small medusae (10–15 mm bell diameter) at a time (Kölliker, 1853).

Nausithoe racemosa is not known to be of conservation concern. Although not well sampled geographically, initial analyses indicate that it is relatively widespread, with species-level differences arising at scales of 2500 km or more (Abboud *et al.*, 2018). Colonies contain many polyps, each capable of polydisc strobilation. Its bipartite life history is, as for other metagenic medusozoans, inferred to confer an ability to persist through periods that are unfavourable for medusae (Arai, 1997: p. 139).

Here, we report a new genome assembly for *Nausithoe racemosa*, the first described genome assembly for a photosymbiotic shallow-water nausithoid, the first for Nausithoidae, and the first for a coronate scyphozoan jellyfish.

Methods

Sample acquisition

Specimens of *Nausithoe racemosa* were picked with long forceps and placed in large plastic bags while diving using SCUBA in Risong Cove, Koror State, Palau (latitude 7.1825, longitude 134.2838) on 2022-07-28. They were transported live to the laboratory (Coral Reef Research Foundation, Koror, Palau) within 1 to 2 hours. At the lab, tissues were biopsied and immediately frozen in liquid nitrogen. Samples were subsequently stored and shipped in an ultra-cold dry shipper, and transferred to a – 80 °C freezer at the University of California, Merced, where they were stored before forwarding to the Wellcome Sanger Institute on dry ice. The specimen used for genome sequencing had specimen ID UCALI0000003 and ToLID jsNauRace1. The jsNauRace1 specimen was also used for RNA sequencing.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of Life Core Laboratory are available on [protocols.io](https://www.protocols.io) (Howard *et al.*, 2025). The jsNauRace1 sample was weighed and [triaged](#) to determine the appropriate extraction protocol. Tissue was homogenised by [powermashing](#) using a Power-Masher II tissue disruptor. HMW DNA was extracted using the [Manual MagAttract v3](#) protocol. DNA was sheared into

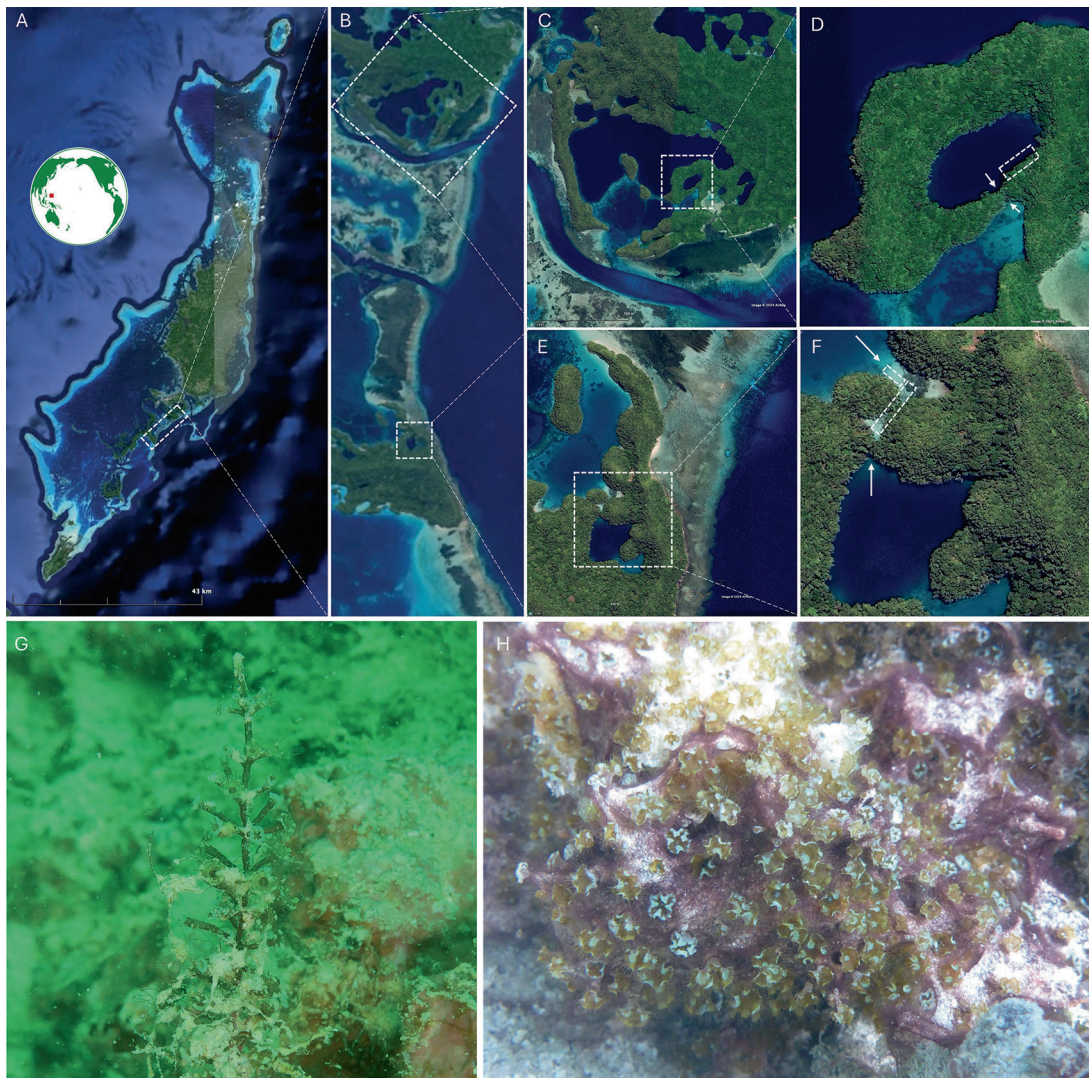


Figure 1. **A.** Palau, with its tropical western Pacific location indicated by the grey square on the inset world map. **B.** The coral reef and channels forming the eastern entrance to Malakal Harbor, framed by the islands of Auluptagel (above) and Urukthapel (below), each of which contains a locality where *Nausithoe racemosa* polyps have been recorded. **C.** Risong Cove, and **D.** detail of Risong Cove; the dashed box shows where *N. racemosa* polyps have regularly been found over the past two decades and where the specimen was sampled for genome sequencing. **E.** Tab Kukau Cove, and **F.** detail of Tab Kukau Cove; the dashed box shows where *N. racemosa* polyps have regularly been found over the past two decades. In **D** and **F**, the opposing arrows indicate shallow channels through which seawater enters and leaves the coves, and which provide small-boat access during mid-to-high tides. **G.** *N. racemosa* polyps at approximately 10 m depth on a rock wall in Risong Cove. **H.** *N. racemosa* polyps along the edge of a shallow channel leading to Tab Kukau Cove. World map from Pixabay. Satellite imagery from Google Earth. Photographs by M. N. Dawson.

an average fragment size of 12–20 kb following the [Megaruptor®3 for LI PacBio](#) protocol. Sheared DNA was purified by [automated SPRI](#) (solid-phase reversible immobilisation). The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from the jsNauRace1 specimen in the Tree of Life Laboratory at the WSI using the [RNA Extraction: Automated MagMax™ mir-Vana](#) protocol. The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

PacBio HiFi library preparation and sequencing

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

The sample was sequenced on a Revio instrument (Pacific Biosciences, California, USA). Prepared libraries with a molarity greater than 2 nM were normalised to 2 nM, and 15 µL was used for making complexes. For libraries with a molarity below 2 nM, the available 10 µL library volume was used without normalisation. Primers were annealed and polymerases were bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2× SMRTbell beads, diluted to the Revio loading concentration of 200–300 pM, and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25 M SMRT cell. SMRT Link software (Pacific Biosciences), a web-based workflow manager, was used to configure and monitor the run and to carry out primary and secondary data analysis.

Hi-C

Sample preparation and crosslinking

The Hi-C sample was prepared from 20–50 mg of frozen tissue of the jsNauRace1 sample using the Arima-HiC v2 kit (Arima Genomics). Following the manufacturer's instructions, tissue was fixed and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. The tissue was homogenised using the Diagnocine Power Masher-II. Crosslinked DNA was digested with a restriction enzyme master mix, biotinylated, and ligated. Clean-up was performed with SPRISelect beads before library preparation. DNA concentration was measured with the Qubit Fluorometer (Thermo Fisher Scientific) and Qubit HS Assay Kit. The biotinylation percentage was estimated using the Arima-HiC v2 QC beads.

Hi-C library preparation and sequencing

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

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

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA was isolated from 100 ng of total RNA using oligo (dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were prepared with an intended fragment size of 100–300 bp. Libraries were quantified, normalised, and pooled equimolarly to a final concentration of 2.8 nM before dilution to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq X, generating paired-end reads.

Genome assembly

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

The HiFi reads were assembled using Hifiasm in Hi-C phasing mode (Cheng *et al.*, 2021, 2022), producing two haplotypes. Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019). Contigs were further scaffolded with Hi-C data 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).

Assembly curation

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. TreeVal was used to generate the flat files and maps for use in curation. Manual curation was conducted primarily in PretextView and HiGlass (Kerpedjiev *et al.*, 2018). Scaffolds were visually inspected and corrected as described by Howe *et al.* (2021). Manual corrections included 175 breaks, 211 joins, and removal of 51 haplotypic duplications. This reduced the scaffold count by 3.2%, reduced the scaffold N50 by 4.5%, and reduced the total assembly length by 0.8%. The curation process is described at [sanger-tol/curation-resources](#). PretextViewSnapshot was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

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

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). PacBio reads were aligned to the assembly using minimap2 (Li, 2018) and processed with SAMtools (Daneczek *et al.*, 2021) to generate coverage tracks. The pipeline runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three basal BUSCO lineages, eukaryota, bacteria and archaea, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND BLASTp (Buchfink *et al.*, 2021). The genome assembly is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database using DIAMOND BLASTx. Sequences without DIAMOND BLASTx hits are extracted using seqtk and aligned to the NT database using BLASTn (Altschul *et al.*, 1990). The outputs are consolidated into a BlobDir for visualisation in BlobToolKit. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Metagenome assembly

The metagenome assembly was generated using MetaMDBG (Benoit *et al.*, 2024). The resulting bin sets of each binning algorithm were optimised and refined using MAGScoT (Rühlemann *et al.*, 2022). PROKKA (Seemann, 2014) was used to identify tRNAs and rRNAs in each bin, CheckM (Parks *et al.*, 2015) (checkM_DB release 2015-01-16) was used to assess bin completeness/contamination, and GTDB-Tk (Chaumeil *et al.*, 2022) (GTDB release 214) was used to taxonomically classify bins. Taxonomic replicate bins were identified using dRep (Olm *et al.*, 2017) with default settings (95% ANI threshold). All bins were assessed for quality and categorised as metagenome-assembled genomes (MAGs) if they met the following criteria: contamination $\leq 5\%$, presence of 5S, 16S, and 23S rRNA genes, at least 18 unique tRNAs, and either $\geq 90\%$ completeness or $\geq 50\%$ completeness with fully circularised chromosomes (Bowers *et al.*, 2017). Bins that did not meet these thresholds, or were identified as taxonomic replicates of MAGs, were retained as 'binned metagenomes' provided they had $\geq 50\%$ completeness and $\leq 10\%$ contamination.

Genome sequence report

Sequence data

PacBio sequencing of the *Nausithoe racemosa* specimen generated 125.95 Gb (gigabases) from 16.42 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 4 540.17 Mb, with a heterozygosity of 0.95% and repeat content of 52.03% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided approximately 26 $\times$ coverage. Hi-C sequencing produced 125.32 Gb from 414.96 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

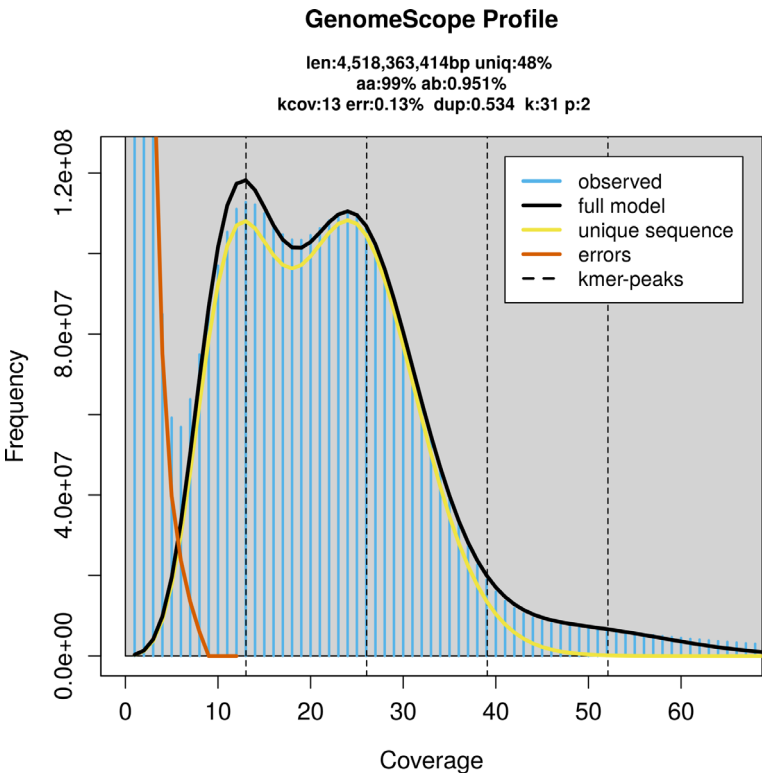


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for *Nausithoe racemosa* (BioProject PRJEB83148).

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	jsNauRace1	jsNauRace1	jsNauRace1
Specimen ID	UCALI0000003	UCALI0000003	UCALI0000003
BioSample (source individual)	SAMEA112358969	SAMEA112358969	SAMEA112358969
BioSample (tissue)	SAMEA112359016	SAMEA112359018	SAMEA112359012
Instrument	Revio	Illumina NovaSeq X	Illumina NovaSeq X
Run accessions	ERR14048091; ERR14048092	ERR14056192	ERR14986724
Read count total	16.42 million reads	414.96 million read pairs	37.58 million read pairs
Base count total	125.95 Gb	125.32 Gb	11.35 Gb

Assembly statistics

The genome was assembled into two haplotypes using Hi-C phasing. Haplotype 1 was curated to chromosome level, while haplotype 2 was assembled to scaffold level. The final assembly has a total length of 4 784.66 Mb in 3 010 scaffolds, with 5 452 gaps, and a scaffold N50 of 224.85 Mb (Table 2).

The scaffold count follows the NCBI assembly statistics and excludes organelle assemblies, which are reported separately where present.

Most of the haplotype 1 assembly sequence (97.34%) was assigned to 20 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3). During genome

Table 2. Genome assembly statistics for *Nausithoe racemosa*.

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	jsNauRace1.hap1.1	jsNauRace1.hap2.1
Assembly accession	GCA_965183845.1	GCA_965183795.1
Assembly level	chromosome	scaffold
Span (Mb)	4 784.66	4 868.20
Number of chromosomes	20	-
Number of contigs	8 462	9 322
Contig N50	1.78 Mb	1.61 Mb
Number of scaffolds	3 010	3 564
Scaffold N50	224.85 Mb	220.4 Mb
Longest scaffold length (Mb)	364.8	366.94
Organelles	Mitochondrial genome: 13.97 kb	-

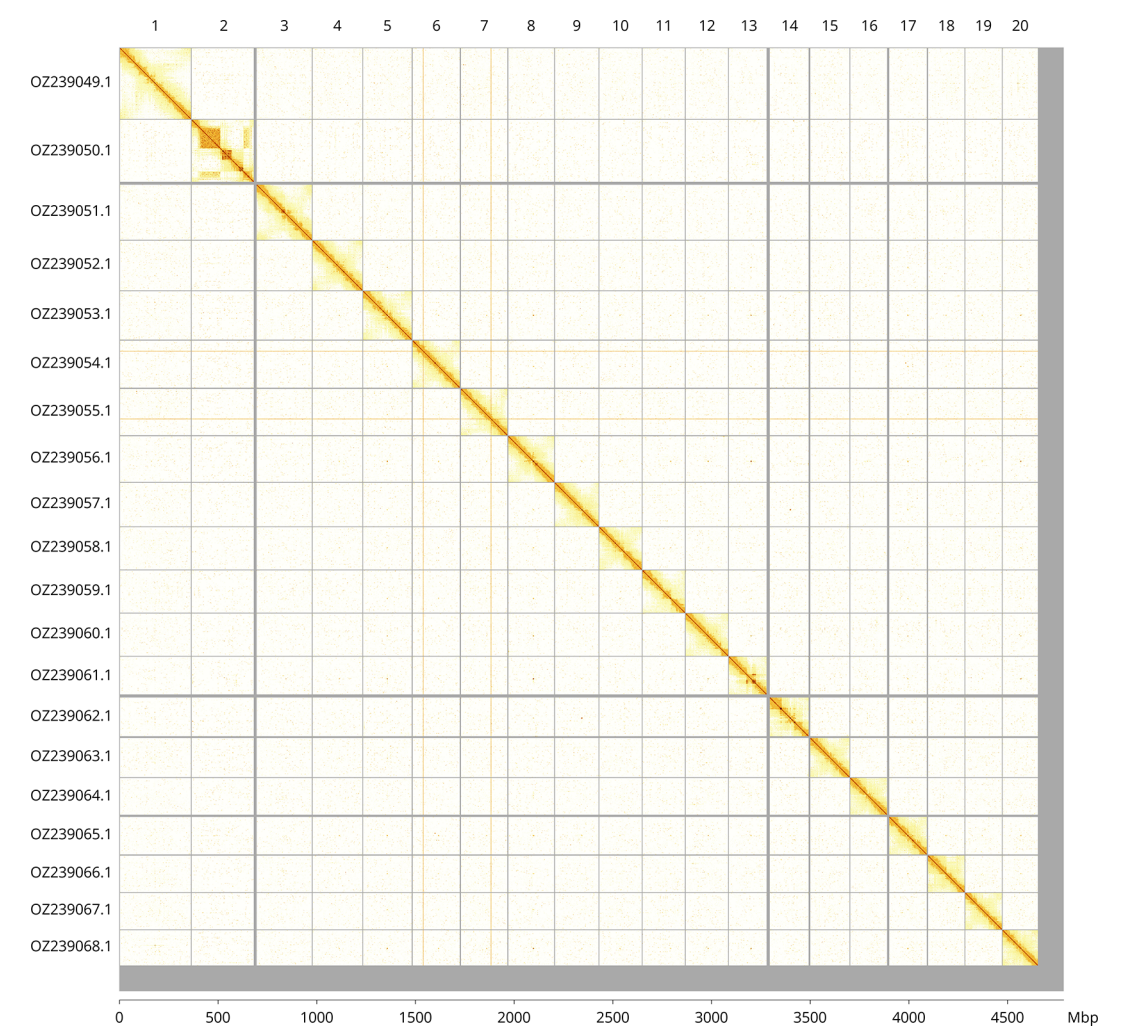


Figure 3. Hi-C contact map of the *Nausithoe racemosa* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the haplotype 1 genome assembly of *Nausithoe racemosa* jsNauRace1.

INSDC accession	Molecule	Length (Mb)	GC%
OZ239049.1	1	364.80	37
OZ239050.1	2	329.92	36.50
OZ239051.1	3	282.21	37
OZ239052.1	4	256.11	37
OZ239053.1	5	251.82	37
OZ239054.1	6	244.29	37
OZ239055.1	7	240.12	37
OZ239056.1	8	235.79	37
OZ239057.1	9	224.85	37
OZ239058.1	10	219.25	37
OZ239059.1	11	218.79	37
OZ239060.1	12	218.73	37
OZ239061.1	13	206.82	36.50
OZ239062.1	14	205.47	37
OZ239063.1	15	202.20	37
OZ239064.1	16	197.93	37
OZ239065.1	17	196.24	37
OZ239066.1	18	189.37	37
OZ239067.1	19	188.17	37
OZ239068.1	20	184.39	37

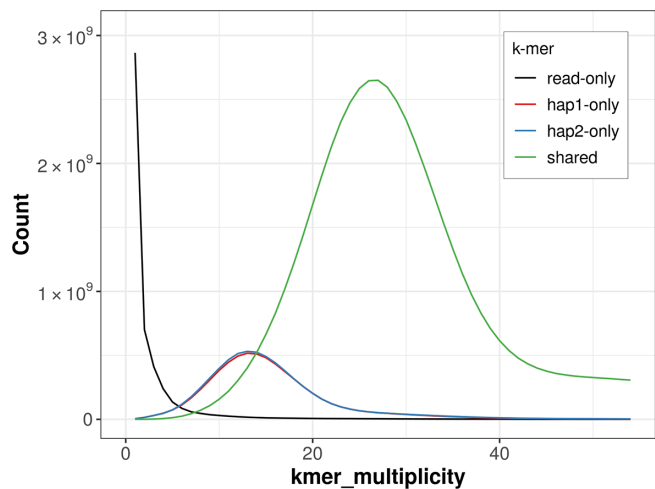


Figure 4. Evaluation of *k*-mer completeness using MerquryFK. This plot illustrates the recovery of *k*-mers from the original read data in the final assemblies. The horizontal axis represents *k*-mer multiplicity, and the vertical axis shows the number of *k*-mers. The black curve represents *k*-mers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

curation, we noted that there are haplotypic inversions on Chromosome 1 (~343.43–356.71 Mb) and Chromosome 9 (~129.07–123.38 Mb). The order and orientation of scaffolds are uncertain in the following regions: Chromosome 2 (~159.59–197.12 Mb) and Chromosome 13 (~128.05–135.50 Mb).

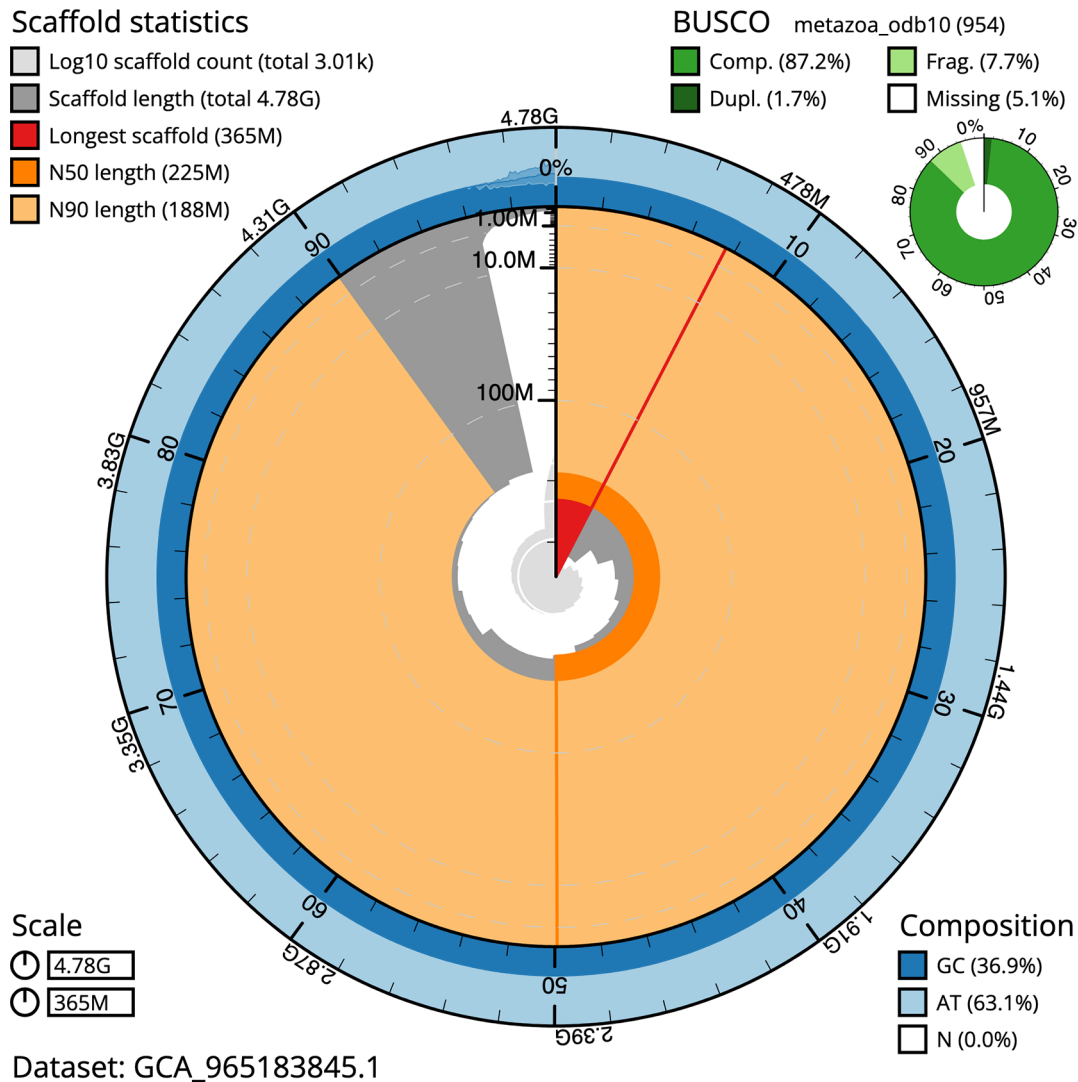


Figure 5. Assembly metrics for jsNauRace1.hap1.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue tracks display the distribution of GC, AT, and N percentages across the bins. Scaffolds are arranged clockwise from longest to shortest and are depicted in dark grey. The longest scaffold is indicated by the red arc, and the deeper orange and pale orange arcs represent the N50 and N90 lengths. A light grey spiral at the centre shows the cumulative scaffold count on a logarithmic scale. A summary of complete, fragmented, duplicated, and missing BUSCO genes in the set is presented at the top right. An interactive version of this figure can be accessed on the [BlobToolKit viewer](#).

The mitochondrial genome was also assembled (length 13.97 kb, OZ239069.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

For haplotype 1, the estimated QV is 60.1, and for haplotype 2, 59.8. When the two haplotypes are combined, the assembly achieves an estimated QV of 59.9. The *k*-mer completeness is 84.19% for haplotype 1, 84.65% for haplotype 2, and 98.91% for the combined haplotypes (Figure 4).

BUSCO analysis using the metazoa_odb10 reference set ($n = 954$) identified 87.2% of the expected gene set (single = 85.5%, duplicated = 1.7%) in haplotype 1. For haplotype 2, BUSCO v.5.7.1 analysis identified 87.7% of the expected gene set (single = 85.7%, duplicated = 2.0%). The snail plot in Figure 5 summarises the scaffold length distribution and other assembly statistics for haplotype 1. The blob plot in Figure 6 shows the distribution of scaffolds by GC proportion and coverage for haplotype 1.

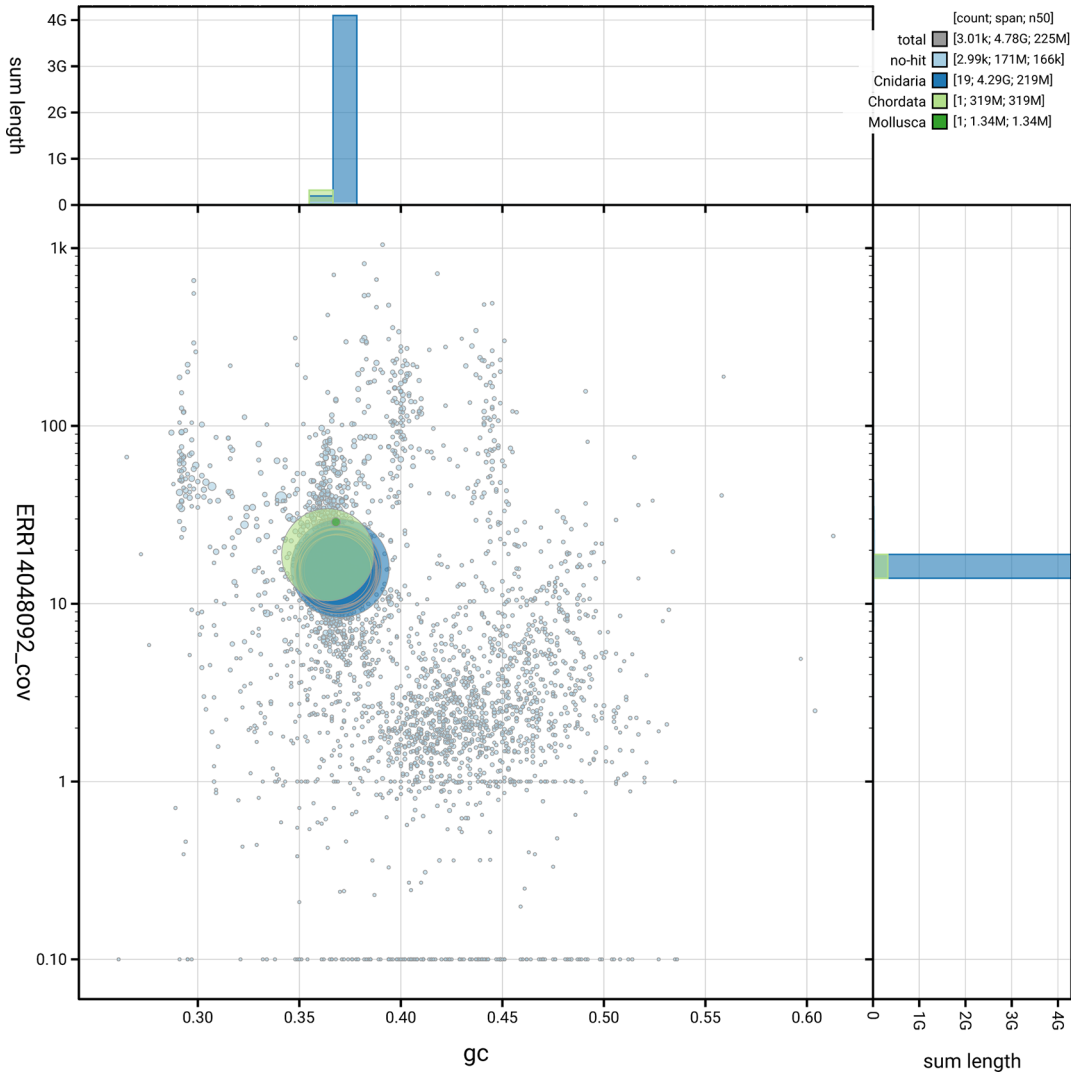


Figure 6. BlobToolKit blob plot for jsNauRace1.hap1.1. The plot shows base coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the [BlobToolKit viewer](#).

Table 4. Earth Biogenome Project summary metrics for the *Nausithoe racemosa* assembly.

Measure	Value	Benchmark
EBP summary (haplotype 1)	6.C.Q60	6.C.Q40
Contig N50 length	1.78 Mb	≥ 1 Mb
Scaffold N50 length	224.85 Mb	= chromosome N50
Consensus quality (QV)	Haplotype 1: 60.1; haplotype 2: 59.8; combined: 59.9	≥ 40
k-mer completeness	Haplotype 1: 84.19%; Haplotype 2: 84.65%; combined: 98.91%	≥ 90%
BUSCO	Haplotype 1: C:87.2% [S:85.5%, D:1.7%], F:7.7%, M:5.1%, n:954; Haplotype 2: C:87.7% [S:85.7%, D:2.0%], F:7.0%, M:5.2%, n:954	S > 90%; D < 5%
Percentage of assembly assigned to chromosomes	97.34%	≥ 90%

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

[Table 4](#) lists the assembly metric benchmarks adapted from [Rhie *et al.* \(2021\)](#) and the Earth BioGenome Project Report on Assembly Standards [January 2026](#). The EBP metric, calculated for haplotype 1, is **6.C.Q60**, meeting the recommended 6.C.Q40 reference standard.

Metagenome report

We recovered one MAG from the metagenome assembly (GCA_965240435.1). It was assigned as *Moorena producens* (NCBI taxid:1155739), with a total genome size of 9.78 Mbp. The completeness was estimated to be 98.7% and the contamination 1.3%.

Author information

Contributors are listed at the following links:

- Members of the [Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team](#)
- Members of [Wellcome Sanger Institute Scientific Operations – Sequencing Operations](#)
- Members of the [Wellcome Sanger Institute Tree of Life Core Informatics team](#)
- Members of the [EBI Aquatic Symbiosis Genomics Data Portal Team](#)
- The [Aquatic Symbiosis Genomics Project leadership](#)

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.

Acknowledgements

Collections in Palau were made under Marine Research Permit RE-22-19 issued by the Ministry of Natural Resources, Environment, and Tourism and under Marine Research/Collection Permit #77 issued by Koror State Government. As ever, many thanks to all at the Coral Reef Research Foundation for outstanding logistical and other support. Thank you also to Karly Higgins-Poling, Laura Martin, Lisa Pageot and Lauren Schiebelhut for help in the field and lab.

Data availability

European Nucleotide Archive: *Nausithoe racemosa*. Accession number [PRJEB83148](#); <https://identifiers.org/ena.embl/PRJEB83148>. The genome sequence is released openly for reuse. The *Nausithoe racemosa* genome sequencing initiative is part of the Aquatic Symbiosis Genomics Project (PRJEB43743) and the Sanger Institute Tree of Life Programme (PRJEB43745). All raw sequence data and the assembly have been deposited in INSDC databases. 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 [Tables 1 and 2](#).

Production code used in genome assembly at the WSI Tree of Life is available at <https://github.com/sanger-tol>. [Table 5](#) lists software versions used in this study.

Table 5. Software versions and sources used for *Nausithoe racemosa*.

Software	Version	Source
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/
BlobToolKit	4.4.5	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.7.1	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
CheckM	2015-01-16	https://ecogenomics.github.io/CheckM/
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
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GTDB-Tk	1.2.1	https://github.com/ECogenomics/GTDBTk
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MAGScoT	1.0.0	https://github.com/ikmb/MAGScoT
MerquryFK	1.1.0-c1	https://github.com/thegenemyers/MERQURY.FK
MetaMDBG	Pre-release	https://github.com/GaetanBenoitDev/metaMDBG
Minimap2	2.24-r1122; 2.28-r1209	https://github.com/lh3/minimap2
MitoHiFi	3.2.2	https://github.com/marcelauliano/MitoHiFi
MultiQC	1.14; 1.17 and 1.18	https://github.com/MultiQC/MultiQC
Nextflow	23.10.0; 24.10.4	https://github.com/nextflow-io/nextflow
PretextSnapshot	0.0.4	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
Prokka	1.14.5	https://github.com/tseemann/prokka
samtools	1.17; 1.18; 1.2; 1.21; 1.6	https://github.com/samtools/samtools
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	v0.7.1	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
Seqtk	1.4-r122	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
TreeVal	1.1.1	https://github.com/sanger-tol/treeval
YaHS	1.2a.2	https://github.com/c-zhou/yahs

References

- Abboud SS, Gómez Daglio L, Dawson MN: **A global estimate of genetic and geographic differentiation in macromedusae—implications for identifying the causes of jellyfish blooms.** *Mar. Ecol. Prog. Ser.* 2018; **591**: 199–216.
[Publisher 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](#)
- Arai MN: *A Functional Biology of Scyphozoa*. London: Chapman & Hall; 1997.
[Publisher Full Text](#)
- Bateman A, Martin M-J, 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](#)
- Bayha KM, Dawson MN, Collins AG, et al.: **Evolutionary relationships among scyphozoan jellyfish families based on complete taxon sampling and phylogenetic analyses of 18S and 28S ribosomal DNA.** *Integr. Comp. Biol.* 2010; **50**(3): 436–455.
[PubMed Abstract](#) | [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**: 1378–1383.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bowers RM, Kyrpides NC, Stepanauskas R, et al.: **Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea.** *Nat. Biotechnol.* 2017; **35**(8): 725–731.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Buchfink B, Reuter K, Drost H-G: **Sensitive protein alignments at tree-of-life scale using DIAMOND**. *Nat. Methods*. 2021; **18**(4): 366–368.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies**. *G3: Genes, Genomes, Genetics*. 2020; **10**(4): 1361–1374.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chaumeil P-A, Mussig AJ, Hugenholtz P, *et al.*: **GTDB-Tk v2: Memory friendly classification with the genome taxonomy database**. *Bioinformatics*. 2022; **38**(23): 5315–5316.
[Publisher 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](#)
- Collins AG: **Phylogeny of Medusozoa and the evolution of cnidarian life cycles**. *J. Evol. Biol.* 2002; **15**(3): 418–432.
[Publisher Full Text](#)
- Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools**. *GigaScience*. 2021; **10**(2): giab008.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Djeghri N, Pondaven P, Stibor H, *et al.*: **Review of the diversity, traits, and ecology of zooxanthellate jellyfishes**. *Mar. Biol.* 2019; **166**: 147.
[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](#)
- Helm RR: **Evolution and development of scyphozoan jellyfish**. *Biol. Rev.* 2018; **93**(2): 1228–1250.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Higgins-Poling KH: **Genetic and behavioral variation in motile marine holobionts (Scyphozoa-Symbiodiniaceae)**. 2024; **77**
[Reference Source](#)
- Hingston MC, Jarms G, Zibrowius H: **The ecology and phylogeny of the Mediterranean Nausithoidae (Cnidaria, Scyphozoa)**. *Vie et Milieu/Life & Environment*. 2007; **57**(1/2): 67–74.
[Reference Source](#)
- Howard C, Denton A, Jackson B, *et al.*: **On the path to reference genomes for all biodiversity: Lessons learned and laboratory protocols created in the Sanger Tree of Life core laboratory over the first 2000 species**. *bioRxiv*. 2025.
[Publisher Full Text](#)
- Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation**. *GigaScience*. 2021; **10**(1): g1aa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jarms G, Båmstedt U, Tiemann H, *et al.*: **The holopelagic life cycle of the deep-sea medusa *Periphylla periphylla* (Scyphozoa, Coronatae)**. *Sarsia*. 1999; **84**(1): 55–65.
[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](#)
- Kölliker A: **Bericht über einige im Herbste 1852 in Messina angestellte vergleichend-anatomische Untersuchungen**. *Z. Wiss. Zool*. 1853; **4**(3–4): 298–370.
[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](#)
- Lajeunesse TC, Parkinson JE, Gabrielson PW, *et al.*: **Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts**. *Curr. Biol.* 2018; **28**(16): 2570–2580.e6.
[PubMed Abstract](#) | [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, Seppely 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.
[Publisher Full Text](#)
- Mianzan HW, Cornelius PFS: **Cubomedusae and Scyphomedusae**. Boltovskoy D, editor. *South Atlantic Zooplankton*. Leiden, The Netherlands: Backhuys Publishers; 1999; Vol. **1**: 513–59.
- Molinari CG, Collins AG, Morandini AC: **A morphological review of the jellyfish genus *Nausithoe* Kölliker, 1853 (Nausithoidea, Coronatae, Scyphozoa, Cnidaria)**. *Zootaxa*. 2023; **5336**(1): 1–32.
[PubMed Abstract](#) | [Publisher Full Text](#)
- OBIS: **Coronatae Vanhöffen, 1892 in Ocean Biodiversity Information System**. 2026.
[Reference Source](#)
- 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): 1043–1055.
[Publisher 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, 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](#)
- Rühlemann MC, Wacker EM, Ellinghaus D, *et al.*: **MAGScoT: A fast, lightweight and accurate bin-refinement tool**. *Bioinformatics*. 2022; **38**(24): 5430–5433.
[Publisher Full Text](#)
- Schoch CL, Ciufo S, Domrachev M, *et al.*: **NCBI taxonomy: A comprehensive update on curation, resources and tools**. *Database*. 2020; **2020**: baaa062.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Seemann T: **Prokka: Rapid prokaryotic genome annotation**. *Bioinformatics*. 2014; **30**(14): 2068–2069.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Trench RK, Luong-van Thinh: ***Gymnodinium linucheae* sp. Nov.: The dinoflagellate symbiont of the jellyfish *Linuche unguiculata***. *Eur. J. Phycol.* 1995; **30**(2): 149–154.
[Publisher Full Text](#)
- Uliano-Silva M, Ferreira JGRN, Krashennikova 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**. 2019 *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE; 2019; 314–24.
[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 03 September 2026

<https://doi.org/10.21956/wellcomeopenres.29980.r163078>

© 2026 Burgess 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.



Shawn M Burgess

National Human Genome Research Institute, Bethesda, USA

The manuscript: *The genome sequence of a coronate scyphozoan jellyfish, Nausithoe racemosa (Komai, 1936) (Coronatae: Nausithoidae), and a metagenome-assembled genome of the associated cyanobacterium Moorena producens* describes the effort to sequence and assemble the genome of a jellyfish that resides in a somewhat uncertain location in the evolutionary tree. The paper is clear and concise on how the genome was generated and uses state of the art approaches to generate a high-quality annotated genome.

The only comment I have regarding the manuscript is that the introduction is a bit hard to follow in terms of how the organism sits in the clade in relation to other closely related jellyfish. I think it would be helpful (but not an absolute requirement for "acceptance") if included in figure 1 there was a tree showing the possible alternative relationships of these jellyfish or the branches in question. Otherwise, I have no other objections.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 31 August 2026

<https://doi.org/10.21956/wellcomeopenres.29980.r163080>

© 2026 Kron N. 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.



Nicholas S Kron

Department of Marine Biology and Ecology, Rosenstiel School of Marine, Atmospheric, and Earth Science, University of Miami, Miami, Florida, USA

Reviewer summary:

This report by Dawson et al. describes the sampling, sequencing, and assembly of a chromosome level genome for the scyphozoan *Nausithoe racemosa* which lacks genomic resources. The authors describe the paucity of sufficient resources in this clade of (Coronatae) to answer questions regarding phylogeny and the evolution of photosymbiosis in the clade. The authors describe the collection of samples from wild caught jellies in Palau and their subsequent sampling and sequencing at the Wellcome Sanger Institute using standard protocols. The genome was assembled from PacBio HiFi reads and Hi-C data with Hifiasm in phasing mode and then scaffolded with Hi-C reads followed by automatd and manual curation with ToL standard pipelines, tools, and workflows. The resulting assembly was at the chromosome scale (20 pseudo-molecules) with a BUSCO completeness at ~87% for metazoa odb10. The author also describe collection of RNAseq reads, assembly of the mitochondrion, and metagenome assembly of 1 symbiont, but do not present analysis or results for these beyond estimated completeness for the one identified MAG that passed quality metrics.

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

Yes. The authors outline a clear paucity of available genetic resources and several specific questions creating this resource could answer.

Are the protocols appropriate and is the work technically sound?

Overall, yes. Protocols described are clear and comprise standard tools and workflows. The estimated 26X coverage is a bit on the low end for high-quality reference genome, but this is to be expected when working with hard to collect non-models.

My one major concern is regarding endosymbionts. The introduction notes that this species is known to carry Symbiodiniaceae as endosymbionts. In other photosymbiotic cnidarians, these can represent a significant fraction of total reads (up to ~1/3 in corals). Is this also the case with this species? Was the life stage collected not generally associated with this symbiont? The authors do not describe decontamination of dinoflagellate reads prior to assembly, so I assume these were included in the initial 26X estimate from GenomeScope and assembled in the primary assembly and purged downstream with ASCC. The authors describe how much manual curation affected assemblies, but do not describe the results of the ASCC run. What fraction of raw reads and/or contigs were purged as contaminants? The final BlobTools plot does suggest successful purging of any symbiont material. This is also relevant to the metagenome assembly, as the MAG methods

appear to target only prokaryote lineages. Given that the primary known endosymbiont is a Eukaryote, the major endosymbiont would be missed. Given that the introduction centers the photosymbiotic nature of this jelly and its specific symbiont, I would like to see this aspect of the assembly addressed (how much symbiont material was recovered, what was done with it, how that impacted the assembly, etc.).

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

For the majority, yes. The tools, versions, and order in which were used are thoroughly documented which is appreciated.

The authors mention in paragraph 1 of 'Genome assembly' that they used 'filtered reads' as input into GenomeScope2 but no prior section describes any form of raw read filtering or QC. Are we to assume this was handled in the SMRT Link primary and secondary data analysis mentioned in the 'PacBio HiFi library preparation and sequencing' section?

Some minor additional notes for this question:

- The authors assess genome completion with metazoa_odb10. This is acceptable for comparison to metazoa genomes at large, but there are more recent and lineage specific BUSCO databases available. While not always strictly followed in the literature, BUSCO best practice is to use the most recent and specific lineage possible, which here would be cnidaria_odb12. Inclusion of BUSCO assessment to this database would provide a better sense of completeness.
- The authors do not specify whether default parameters were used. Although this is generally assumed when no specification is provided, it would be good to specify that default params were used and why.
- The authors also do not specify the versions/tags of Docker containers used which would also be appreciated for reproducibility although not strictly necessary.

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

Yes. The author list exact accessions for the raw data and assemblies in public repositories.

Additional notes

- The authors describe the collection of RNA and sequencing of said RNA. While this report is a genome assembly report, not a full resource report with transcriptome and annotation, it would be nice to see the fraction of read alignment to the final assembly as another metric of completeness.
- The authors note a mitochondrial assembly with MitoHiFi. They report it is included in the final assembly and as a stand-alone record. I am not familiar with the Wellcome Sanger Institute standard for reporting genomes (e.g. if the mitochondrial assembly is customarily not discussed since it is not part of the nuclear genome) but given the mitochondrial assembly is reported it would be nice to at least see the MitoHiFi report and annotation results (especially since there is a metagenome report as well).

Overall assessment

The authors present a very thorough report and generally high-quality assembly for this under sampled clade of marine organisms. This resource will undoubtedly be very useful. Barring some

questions regarding the Symbiodiniaceae endosymbiont, this report is ready for indexing. Well done authors.

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

Yes

Are the protocols appropriate and is the work technically sound?

Partly

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

Partly

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: genome assembly and annotation, transcriptomics

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

Reviewer Report 06 August 2026

<https://doi.org/10.21956/wellcomeopenres.29980.r163075>

© 2026 Xing B. 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.



Bingpeng Xing

Third Institute of Oceanography Ministry of Natural Resources, Xiamen, Fujian, China

The authors report a haplotype-resolved, chromosome-level genome assembly of the coronate scyphozoan *Nausithoe racemosa*. The two haplotypes are approximately 4.78 and 4.87 Gb in size, respectively, with 97.34% of haplotype 1 anchored to 20 chromosome-level pseudomolecules. The assembly QV is close to 60, and the combined k-mer completeness of the two haplotypes is 98.91%. These data represent a valuable genomic resource for research on Coronatae, the family Nausithoidae, and comparative genomics of cnidarians.

The principal concerns with the current manuscript are as follows:

The evidence supporting host species identification and the information on voucher specimens are insufficient. It is unclear whether the identification was based on the medusa stage, the polyp stage, or life-history stages obtained through laboratory culture. The manuscript also does not state whether molecular identification using markers such as COI, 16S, 18S, or 28S was performed. The description of the cyanobacterium as “associated” in the title is not supported by direct

evidence.

Table 1 lists only the host PacBio HiFi, Hi-C, and RNA-seq datasets, without identifying a separate dataset corresponding to the reported "metagenome data."

There is an apparent inconsistency between the reported GTDB-Tk software version and the GTDB database release.

The description and quality assessment of the MAG are far less comprehensive than would be expected for a dataset that constitutes a major component of the article title.

Finally, the presentation of the host genome quality is somewhat selective. The manuscript does not adequately explain why the BUSCO completeness and haplotype-specific k-mer completeness values fail to meet the benchmarks listed in the corresponding table.

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

Partly

Are the protocols appropriate and is the work technically sound?

Partly

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

Partly

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

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: genome, DNA barcoding, mitogenome, marine biodiversity

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