



DATA NOTE

The chromosomal genome sequence of the carnivorous sponge, *Lycopodina hypogea* (Vacelet & Boury-Esnault, 1996) (Poecilosclerida: Cladorhizidae) and its associated microbial metagenome sequences

[version 1; peer review: 4 approved]

Thierry Pérez¹, Jean Vacelet¹, Dirk Erpenbeck^{id}², Ute Hentschel^{id}³, Graeme Oatley^{id}⁴, Elizabeth Sinclair^{id}⁴, Eerik Aunin⁴, Noah Gettle^{id}⁴, Camilla Santos⁴, Michael Paulini⁴, Haoyu Niu⁴, Victoria McKenna⁴, Rebecca O'Brien^{id}⁴,

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

¹Mediterranean Institute of Biodiversity and Ecology (IMBE), Station Marine d'Endoume, Aix-Marseille University, Avignon University, IRD, Marseille, France

²Department of Earth and Environmental Sciences, Palaeontology & Geobiology, LMU München, Munich, Germany

³GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

⁴Tree of Life Programme, Wellcome Sanger Institute, Hinxton, England, UK

V1 First published: 17 Feb 2026, 11:130
<https://doi.org/10.12688/wellcomeopenres.25959.1>

Latest published: 17 Feb 2026, 11:130
<https://doi.org/10.12688/wellcomeopenres.25959.1>

Abstract

We present a genome assembly from an individual *Lycopodina hypogea* (carnivorous sponge; Porifera; Demospongiae; Poecilosclerida; Cladorhizidae). The genome sequence has a total length of 235.10 megabases. Most of the assembly (98.85%) is scaffolded into 15 chromosomal pseudomolecules. The mitochondrial genome has also been assembled, with a length of 31.1 kilobases. Gene annotation of this assembly by Ensembl identified 16 317 protein-coding genes. From the metagenome data we recovered 39 bins, of which 27 were high-quality MAGs, including four fully circularised genomes. The MAGs included archaea and bacteria

Open Peer Review

Approval Status

	1	2	3	4
version 1				
17 Feb 2026	view	view	view	view

1. **Emily C Giles** ^{id}, Cawthron Institute (Ringgold ID: 5732), Nelson, New Zealand
2. **Poppy Hesketh-Best** ^{id}, IGTP Institut Germans Trias i Pujol, Barcelona, Spain

involved in nitrification and sulfate-reduction as well as known sponge symbionts affiliated with Gammaproteobacteria (*Candidatus* Spongihabitans, Porisulfidus) and Acidimicrobiales (*Candidatus* Poriferisodalaceae), among others.

Keywords

Lycopodina hypogea; carnivorous sponge; genome sequence; chromosomal; Poecilosclerida; microbial metagenome assembly



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

3. **Estelle Proux-Wéra** , Stockholm University, stockholm, Sweden

4. **Jayan Duminda M Senevirathna** , Uva Wellassa University, Badulla, Sri Lanka

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: Pérez T: Resources, Writing – Original Draft Preparation, Writing – Review & Editing; Vacelet J: Writing – Original Draft Preparation, Writing – Review & Editing; Erpenbeck D: Investigation; Hentschel U: Funding Acquisition, Project Administration, Supervision, 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).

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

Copyright: © 2026 Pérez T *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: Pérez T, Vacelet J, Erpenbeck D *et al.* **The chromosomal genome sequence of the carnivorous sponge, *Lycopodina hypogea* (Vacelet & Boury-Esnault, 1996) (Poecilosclerida: Cladorhizidae) and its associated microbial metagenome sequences [version 1; peer review: 4 approved]** Wellcome Open Research 2026, 11:130 <https://doi.org/10.12688/wellcomeopenres.25959.1>

First published: 17 Feb 2026, 11:130 <https://doi.org/10.12688/wellcomeopenres.25959.1>

Species taxonomy

Eukaryota; Opisthokonta; Metazoa; Porifera; Demospongiae; Heteroscleromorpha; Poecilosclerida; Cladorhizidae; *Lycopodina*; *Lycopodina hypogea* (Vacelet & Boury-Esnault, 1996) (NCBI:txid2491183)

Background

Lycopodina hypogea is one of the most iconic sponges described in the last 30 years (Vacelet & Boury-Esnault, 1996). The discovery of this species redefined the concept of sponges, as this sponge has replaced its filter-feeding body plan with a carnivorous feeding habit. This marine sponge belongs to Cladorhizidae, a family of Porifera generally living in deep-sea environments and mostly sharing this unusual feeding strategy.

The genus *Lycopodina* includes the deepest known sponge ever found, *L. occidentalis* (Lambe, 1883), which was reported by Koltun (1970) at a depth of 8 840 m. *L. hypogea* has a Mediterranean and south European Atlantic shelf distribution, in bathyal environments and in dark littoral caves (Chevaldonné *et al.*, 2015). Thus, this sponge can be easily collected by scuba diving and maintained and cultured under laboratory conditions for years. Sub-populations of tens of specimens only require rather small aquaria, kept at a stable temperature (13–15 °C), with a monthly food supply (*Artemia* nauplii) and a supply of fresh seawater. This model sponge and its very simple culture methods thus offer exceptional experimental conditions, which have already led to significant advances in our knowledge of sponges in this family and of deep-sea organisms in general.

Lycopodina hypogea is small, only 20 mm high, composed of a fixation basis, a thin peduncle and an ovoid body bearing numerous long filaments. All these structures are maintained by monaxial siliceous spicules. The filaments are covered in small, hook-shaped spicules (anisochelae) disposed at right angles to the surface and giving an adhesiveness ensuring the capture of “hairy” prey. The sponge body comprises diverse cellular types, pinacocytes, archaeocytes, sclerocytes, cells with inclusions and two types of bacteriocytes, well-illustrated in transmission electron microscopy (Vacelet & Boury-Esnault, 1996).

Previous DNA sequencing using 454 pyrosequencing enabled the discovery of a highly diverse microbial community, comprising at least 22 prokaryotic phyla, dominated by Proteobacteria, Bacteroidetes and Thaumarchaeota. A high abundance of ammonia-oxidising archaea and a dominance of sulphate oxidising/reducing bacteria were observed (Dupont *et al.*, 2013; Dupont *et al.*, 2014). Few other functional roles of the sponge microbiome were investigated through antioxidant, antimicrobial and chitinase assays (Dupont *et al.*, 2013). Moreover, methanotrophic symbiotic bacteria may also exist, as they were found in other Cladorhizidae (Vacelet *et al.*, 1995; Vacelet *et al.*, 1996).

Sexual reproduction could also be investigated under experimental conditions. In the absence of choanocytes, from which male gametes are derived in the vast majority of known

sponges, spermatozoa here originate from archaeocyte-like cells. They are concentrated and then transported in special spermatophores, harbouring special spicules (forceps) which, once the spermatophore is released by the filament, ensure its capture on filaments of an individual to be fertilised (Vacelet *et al.*, 2022). Although *L. hypogea* regularly reproduces in aquaria, its oogenesis and larval development remain enigmatic.

After the capture of a prey item, generally small crustaceans, polychaetes or other invertebrates with appendages that can become entangled in the filaments of the sponge, the digestive process starts by migration of sponge cells towards the prey. These cells, including cells of the body, but also of filaments not involved in the capture or even cells of the fixation base, cover the prey and incorporate fragments of its cells by phagocytosis. Such a mechanism of digestion is highly unusual in animals, and the movement and high rate of cell renewal make this sponge a model of choice in biology, particularly for studying processes such as apoptosis (Baghdiguian *et al.*, 2023; Godefroy *et al.*, 2019; Le Goff *et al.*, 2022).

So far, no toxic or paralysing secretion, and no digestive cavity or digestive enzyme, has been identified. However, chlorinated compounds have been detected in the water around *L. hypogea* kept in aquaria, and a hypothesis currently being studied would give this carnivorous sponge the ability to produce specialised metabolites that could both attract prey and contribute to their immobilisation (Pérez and coll., in progress).

These examples of fundamental knowledge acquired thanks to the accessibility and ease of cultivation of these sponges illustrate how access to the entire hologenome of this model sponge will make it possible to address many specific scientific questions on the biology and ecology of carnivorous sponges and other deep-sea organisms, as well as more general questions on the biology of metazoans.

We present a chromosome-level genome sequence for *Lycopodina hypogea*. The assembly was generated as part of the Aquatic Symbiosis Genomics project, using the Tree of Life pipeline from a specimen collected from La Ciotat, France (Figure 1).

Methods

Sample acquisition

The specimen used for genome sequencing was an adult *Lycopodina hypogea* (specimen ID GHC0000171, ToLID odLycHypo2; Figure 1). A second specimen was used for Hi-C sequencing (specimen ID GHC0000173, ToLID odLycHypo4) and a third for RNA sequencing (specimen ID GHC0000177, ToLID odLycHypo8). The samples were collected from La Ciotat, France (latitude 43.1633, longitude 5.5999) on 2021-03-23. Thierry Pérez collected and formally identified the species.

Nucleic acid extraction

Protocols for high molecular weight (HMW) DNA extraction developed at the Wellcome Sanger Institute (WSI) Tree of



Figure 1. Image of the *Lycopodium hypogaea* (odLycHypo2) specimen used for genome sequencing (photograph by Thierry Pérez).

Life Core Laboratory are available on protocols.io (Howard *et al.*, 2025). The odLycHypo2 sample was weighed and triaged to determine the appropriate extraction protocol. Tissue was homogenised using the [sponge squeezing](#) protocol. HMW DNA was extracted using the [Manual MagAttract v3](#) protocol. We used centrifuge-mediated fragmentation to produce DNA fragments in the 8–10 kb range, following the [Covaris g-TUBE](#) protocol for ultra-low input (ULI). 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. For this sample, the final post-shearing DNA had a Qubit concentration of 1.9 ng/μL and a yield of 741.00 ng.

RNA was extracted from whole organism tissue of odLycHypo8 in the Tree of Life Laboratory at the WSI using the [RNA Extraction: Automated MagMax™ mirVana 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. Prior to library preparation, the DNA was fragmented to ~10 kb. Ultra-low-input (ULI) libraries were prepared using the PacBio SMRTbell® Express Template Prep Kit 2.0 and gDNA Sample Amplification Kit. Samples were normalised to 20 ng DNA. Single-strand overhang removal, DNA damage repair, and end-repair/A-tailing were performed according to the manufacturer's instructions, followed by adapter ligation. A 0.85× pre-PCR clean-up was carried out with Promega ProNex beads.

The DNA was evenly divided into two aliquots for dual PCR (reactions A and B), both following the manufacturer's protocol. A 0.85× post-PCR clean-up was performed with ProNex beads. DNA concentration was measured using a Qubit Fluorometer v4.0 (Thermo Fisher Scientific) with the Qubit HS Assay Kit, and fragment size was assessed on an Agilent Femto Pulse Automated Pulsed Field CE Instrument (Agilent Technologies) using the gDNA 55 kb BAC analysis kit. PCR reactions A and B were then pooled, ensuring a total mass of ≥500 ng in 47.4 μL.

The pooled sample underwent another round of DNA damage repair, end-repair/A-tailing, and hairpin adapter ligation. A 1× clean-up was performed with ProNex beads, followed by DNA quantification using the Qubit and fragment size analysis using the Agilent Femto Pulse. Size selection was performed on the Sage Sciences PippinHT system, with target fragment size determined by Femto Pulse analysis (typically 4–9 kb). Size-selected libraries were cleaned with 1.0× ProNex beads and normalised to 2 nM before sequencing.

The sample was sequenced on a Revio instrument (Pacific Biosciences). The prepared library was normalised to 2 nM, and 15 μL was used for making complexes. Primers were annealed and polymerases bound to generate circularised complexes, following the manufacturer's instructions. Complexes were purified using 1.2X SMRTbell beads, then diluted to the Revio loading concentration (200–300 pM) and spiked with a Revio sequencing internal control. The sample was sequenced on a Revio 25M SMRT cell. The 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 from the odLycHypo4 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 Diagenode 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 6000.

RNA library preparation and sequencing

Libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. Poly(A) mRNA in the total RNA solution was isolated using oligo(dT) beads, converted to cDNA, and uniquely indexed; 14 PCR cycles were performed. Libraries were size-selected to produce fragments between 100–300 bp. Libraries were quantified, normalised, pooled to a final concentration of 2.8 nM, and diluted to 150 pM for loading. Sequencing was carried out on the Illumina NovaSeq 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 (Cheng *et al.*, 2021) with the --primary option. Haplotypic duplications were identified and removed using purge_dups (Guan *et al.*, 2020). The Hi-C reads (Rao *et al.*, 2014) were mapped to the primary contigs using bwa-mem2 (Vasimuddin *et al.*, 2019), and the contigs were scaffolded in YaHS (Zhou *et al.*, 2023) with the --break option for handling potential misassemblies. The scaffolded assemblies were evaluated using Gfastats (Formenti *et al.*, 2022), BUSCO (Manni *et al.*, 2021) and MERQURY.FK (Rhie *et al.*, 2020).

The mitochondrial genome was assembled using 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 49 breaks and 19 joins. This reduced the scaffold count by 89.0%, increased the scaffold N50 by 15.6%, and reduced the total assembly length by 23.9%. The curation process is described at <https://gitlab.com/wtsi-grit/rapid-curation>. PretextView was used to generate a Hi-C contact map of the final assembly.

Assembly quality assessment

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

The genome was analysed using the BlobToolKit pipeline, a Nextflow implementation of the earlier Snakemake version (Challis *et al.*, 2020). The pipeline aligns PacBio reads using minimap2 (Li, 2018) and SAMtools (Danecek *et al.*, 2021) to generate coverage tracks. It runs BUSCO (Manni *et al.*, 2021) using lineages identified from the NCBI Taxonomy (Schoch *et al.*, 2020). For the three domain-level lineages, BUSCO genes are aligned to the UniProt Reference Proteomes database (Bateman *et al.*, 2023) using DIAMOND blastp (Buchfink *et al.*, 2021). The genome is divided into chunks based on the density of BUSCO genes from the closest taxonomic lineage, and each chunk is aligned to the UniProt Reference Proteomes database with DIAMOND blastx. Sequences without hits are chunked using seqtk and aligned to the NT database with blastn (Altschul *et al.*, 1990). The BlobToolKit suite consolidates all outputs into a blobdir for visualisation. The BlobToolKit pipeline was developed using nf-core tooling (Ewels *et al.*, 2020) and MultiQC (Ewels *et al.*, 2016), with containerisation through Docker (Merkel, 2014) and Singularity (Kurtzer *et al.*, 2017).

Metagenome assembly

The metagenome assembly was generated using MetaMDBG (Benoit *et al.*, 2024) and binned using dastool_raw. 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 ≤ 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 (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 ≥ 50% completeness and ≤ 10% contamination. A taxonomic tree of the bins was constructed from NCBI classifications using ete3 (Huerta-Cepas *et al.*, 2016) and visualised with matplotlib.

Genome sequence report

Sequence data

PacBio sequencing of the *Lycopodium hypogea* specimen generated 138.64 Gb (gigabases) from 17.91 million reads, which were used to assemble the genome. GenomeScope2.0 analysis estimated the haploid genome size at 788.73 Mb, with a heterozygosity of 1.55% and repeat content of 81.18% (Figure 2). These estimates guided expectations for the assembly. Based on the estimated genome size, the sequencing data provided

approximately 77× coverage. Hi-C sequencing produced 125.18 Gb from 828.98 million reads, which were used to scaffold the assembly. RNA sequencing data were also generated and are available in public sequence repositories. Table 1 summarises the specimen and sequencing details.

Assembly statistics

The primary haplotype was assembled, and contigs corresponding to an alternate haplotype were also deposited in INSDC

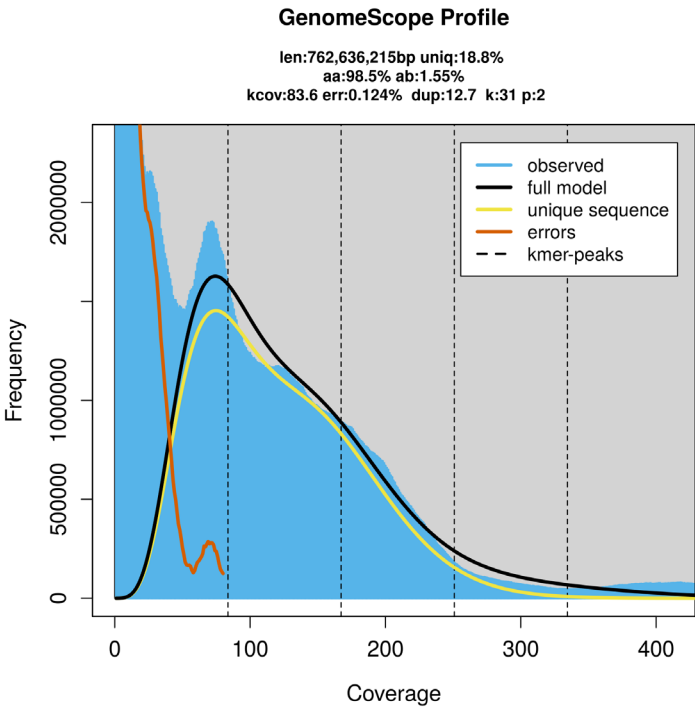


Figure 2. Frequency distribution of *k*-mers generated using GenomeScope2. The plot shows observed and modelled *k*-mer spectra, providing estimates of genome size, heterozygosity, and repeat content based on unassembled sequencing reads.

Table 1. Specimen and sequencing data for BioProject PRJEB72483.

Platform	PacBio HiFi	Hi-C	RNA-seq
ToLID	odLychHypo2	odLychHypo4	odLychHypo8
Specimen ID	GHC0000171	GHC0000173	GHC0000177
BioSample (source individual)	SAMEA9463981	SAMEA9463983	SAMEA9463987
BioSample (tissue)	SAMEA9463992	SAMEA9463994	SAMEA9463998
Tissue	whole organism	whole organism	whole organism
Instrument	Revio	Illumina NovaSeq 6000	Illumina NovaSeq X
Run accessions	ERR12666653; ERR14209102	ERR12668765	ERR13669965
Read count total	17.91 million	828.98 million	102.06 million
Base count total	138.64 Gb	125.18 Gb	15.41 Gb

databases. The final assembly has a total length of 235.10 Mb in 199 scaffolds, with 1 930 gaps, and a scaffold N50 of 16.02 Mb (Table 2).

Most of the assembly sequence (98.85%) was assigned to 15 chromosomal-level scaffolds. These chromosome-level scaffolds, confirmed by Hi-C data, are named according to size (Figure 3; Table 3).

The mitochondrial genome was also assembled (length 31.1 kb, OZ017794.1). This sequence is included as a contig in the multifasta file of the genome submission and as a standalone record.

Assembly quality metrics

BUSCO v.5.5.0 analysis using the metazoa_odb10 reference set ($n = 954$) identified 69.0% of the expected gene set (single = 68.3%, duplicated = 0.6%). The snail plot in Figure 4 summarises the scaffold length distribution and other assembly statistics for the primary assembly. The blob plot in

Table 2. Genome assembly statistics.	
Assembly name	odLycHypo2.1
Assembly accession	GCA_963969325.1
Alternate haplotype accession	GCA_963969335.1
Assembly level	chromosome
Span (Mb)	235.10
Number of chromosomes	15
Number of contigs	2 129
Contig N50	0.17 Mb
Number of scaffolds	199
Scaffold N50	16.02 Mb
Organelles	Mitochondrion: 31.1 kb
BUSCO (using the metazoa_odb10 reference)	C:69.0% [S:68.3%; D:0.6%]; F:11.6%; M:19.4%; n:954

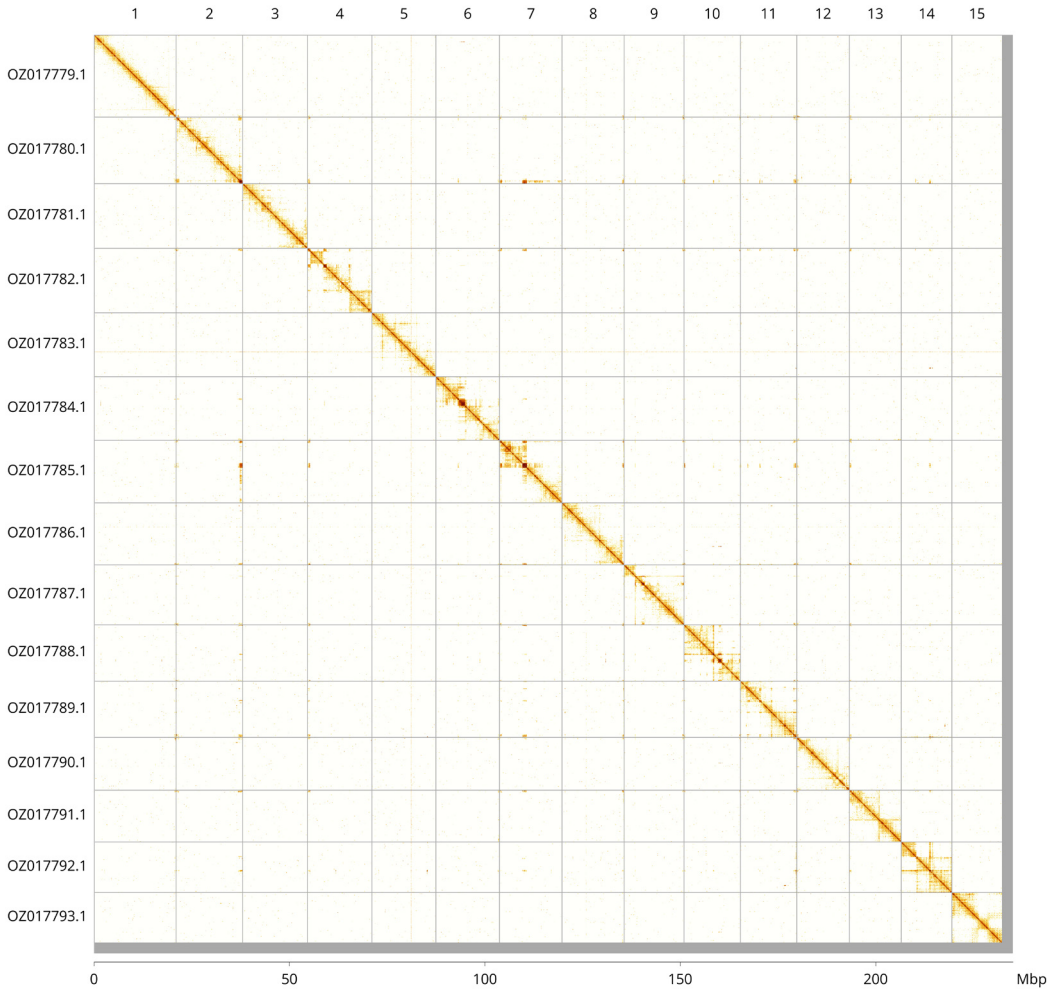


Figure 3. Hi-C contact map of the *Lycopodina hypogaea* genome assembly. Assembled chromosomes are shown in order of size and labelled along the axes, with a megabase scale shown below. The plot was generated using PretextSnapshot.

Table 3. Chromosomal pseudomolecules in the primary genome assembly of *Lycopodia hypogaea* odLycHypo2.

INSDC accession	Molecule	Length (Mb)	GC%
OZ017779.1	1	21.07	39
OZ017780.1	2	16.99	39
OZ017781.1	3	16.60	39.50
OZ017782.1	4	16.47	39
OZ017783.1	5	16.37	39
OZ017784.1	6	16.29	39

INSDC accession	Molecule	Length (Mb)	GC%
OZ017785.1	7	16.02	38.50
OZ017786.1	8	15.84	38.50
OZ017787.1	9	15.31	39.50
OZ017788.1	10	14.45	38
OZ017789.1	11	14.39	38.50
OZ017790.1	12	13.51	38.50
OZ017791.1	13	13.26	39
OZ017792.1	14	12.94	38.50
OZ017793.1	15	12.89	39.50

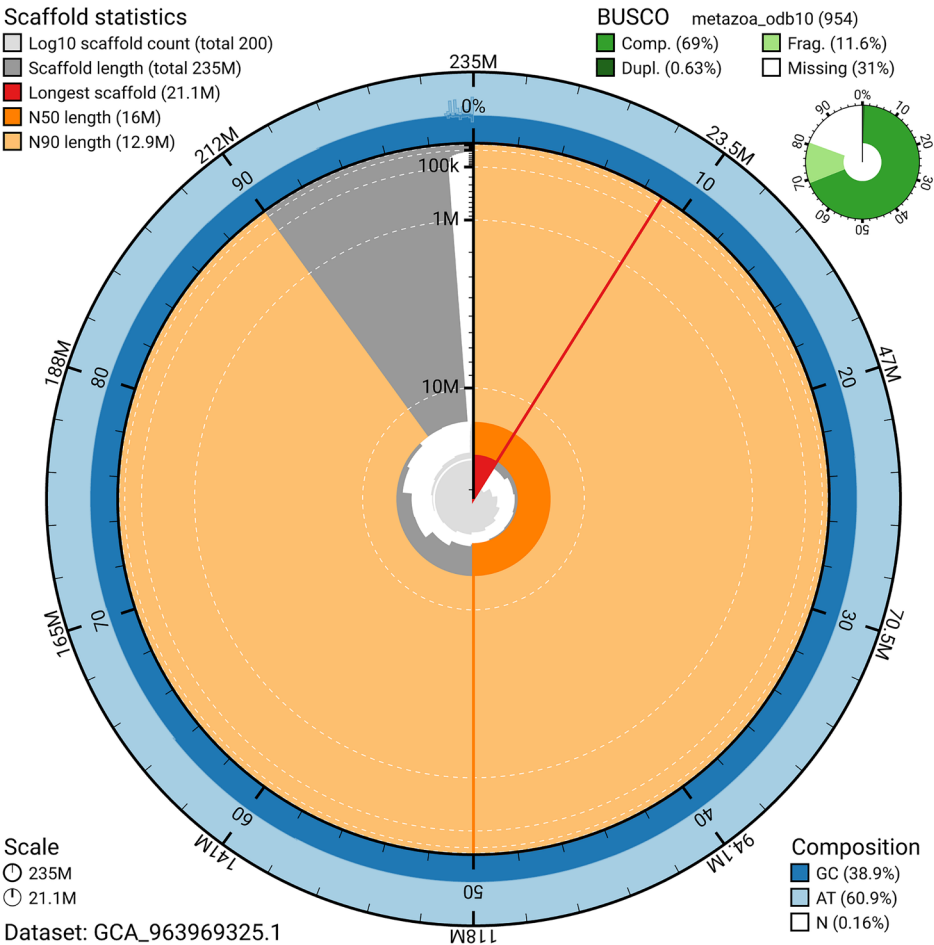


Figure 4. Assembly metrics for odLycHypo2.1. The BlobToolKit snail plot provides an overview of assembly metrics and BUSCO gene completeness. The circumference represents the length of the whole genome sequence, and the main plot is divided into 1 000 bins around the circumference. The outermost blue 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 metazoa_odb10 set is presented at the top right. An interactive version of this figure can be accessed on the [BlobToolKit viewer](#).

Figure 5 shows the distribution of scaffolds by GC proportion and coverage.

Taxonomic verification

The CO1 barcoding sequence (Folmer fragment) of this specimen is 100% identical to *L. hypogea* sequences HG424317–HG424322. These sequences were published in a population study on *L. hypogea* (Chevaldonné *et al.*, 2015).

Genome annotation report

The *Lycopodia hypogea* genome assembly (GCA_963969325.1) was annotated by Ensembl at the European Bioinformatics Institute (EBI). This annotation includes 16 357 transcribed mRNAs from 16 317 protein-coding genes. The average transcript length is 5 381.60 bp, with an average of 5.36 exons per transcript. For further information about the annotation, please refer to the [Ensembl annotation page](#).

Metagenome report

We recovered 39 bins from the metagenome assembly (Figure 6), of which 27 met the criteria for MAGs, including four fully circularised genomes. The recovered bins represented 12 bacterial and archaeal phyla, with genome sizes ranging from 0.87 to 7.67 Mbp (mean: 2.51 ± 1.36 Mbp). Mean completeness was 88.2% ($\pm 13.5\%$) with 1.4% ($\pm 1.7\%$) contamination. Figure 7 summarises the taxa and quality of the metagenome bins. The full per-bin table of taxa and quality metrics for the metagenome bins is available on [Zenodo](#).

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

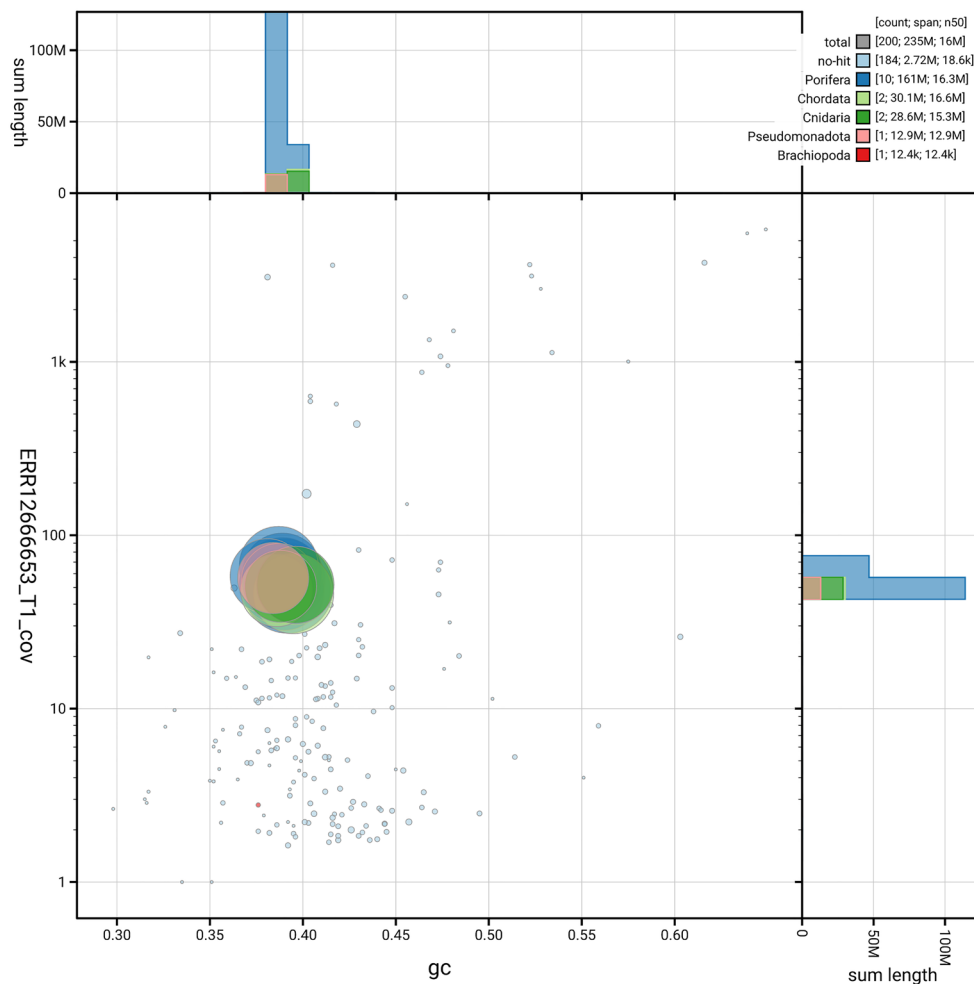


Figure 5. BlobToolKit GC-coverage plot for odLychHypo2.1. Blob plot showing sequence coverage (vertical axis) and GC content (horizontal axis). The circles represent scaffolds, with the size proportional to scaffold length and the colour representing phylum membership. The histograms along the axes display the total length of sequences distributed across different levels of coverage and GC content. An interactive version of this figure is available on the [BlobToolKit viewer](#).

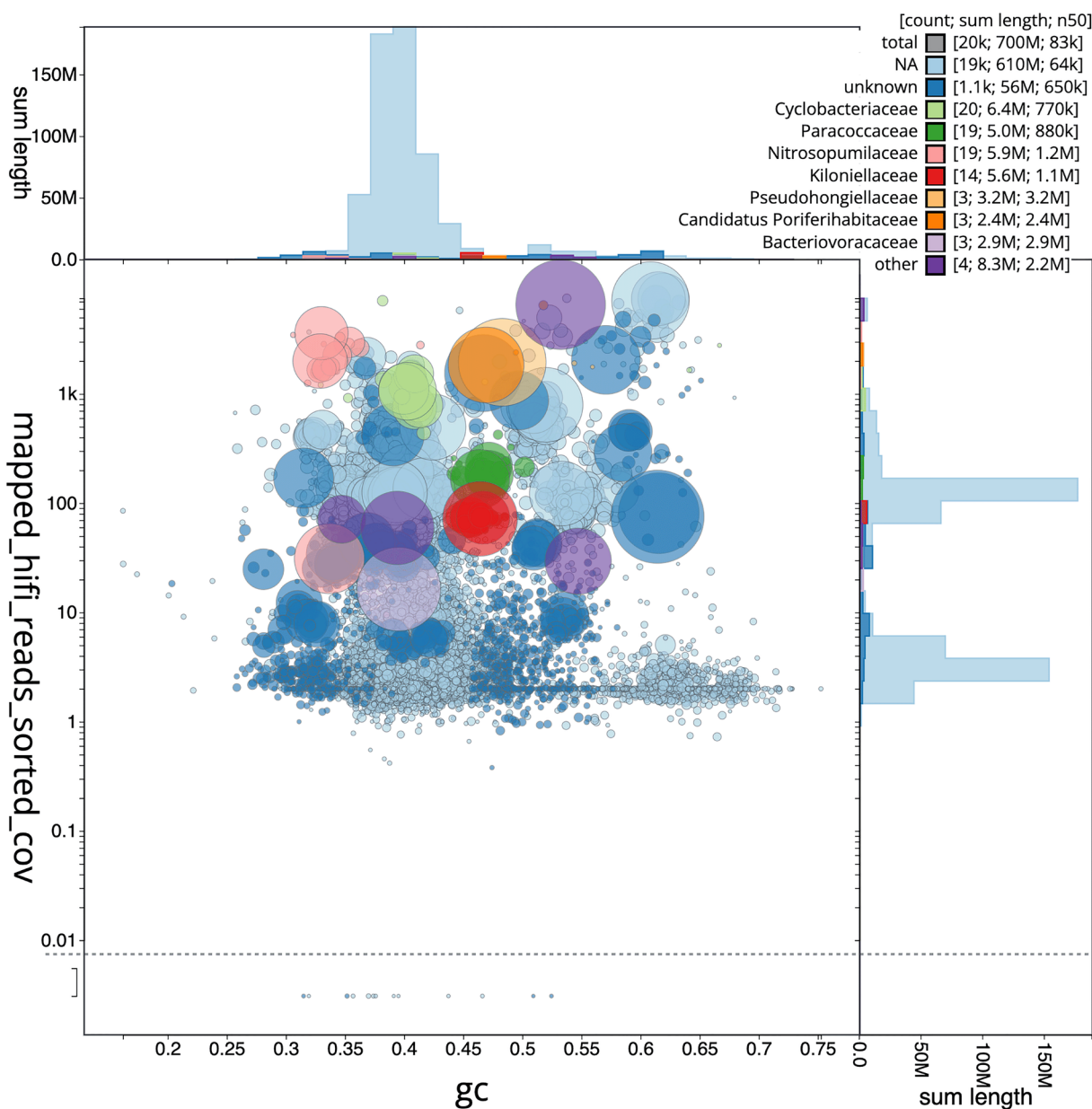


Figure 6. Blob plot of base coverage mapped against GC proportion for sequences in the *Lycopodium hypogaea* metagenome. Binned contigs are coloured by family. Circles are sized in proportion to sequence length on a square-root scale, ranging from 510 to 6 879 066. Histograms show the distribution of sequence length sum along each axis. An interactive version of this figure may be viewed [here](#).

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.

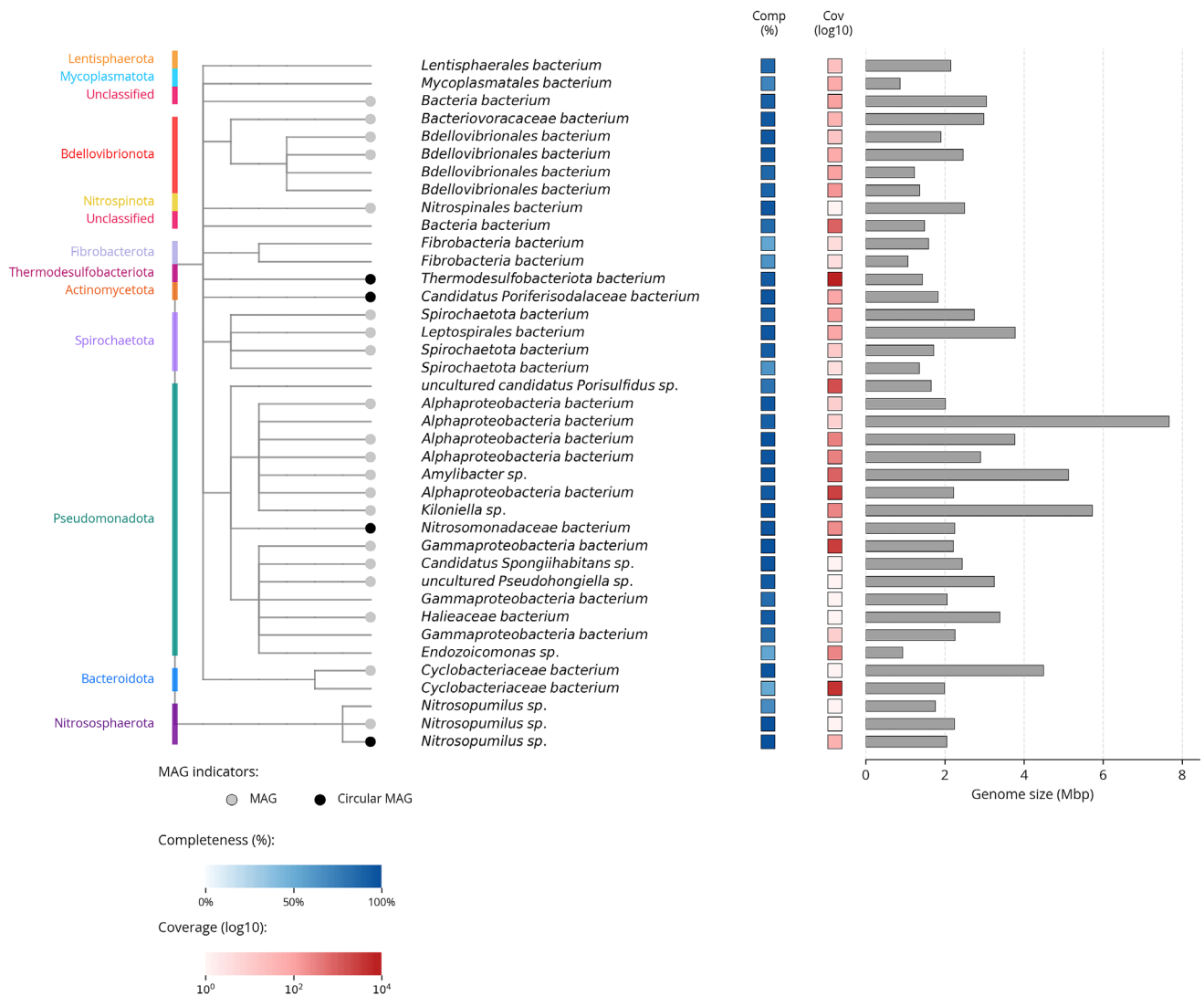


Figure 7. Taxonomic tree based on taxonomic classifications of metagenome bins, constructed using ete3. Colours indicate phylum-level taxonomy. Tracks show genome completeness (blue), sequencing coverage (red, $\log_{10}$), and genome size (grey bars, Mbp). High-quality MAGs are marked with grey circles; fully circularised MAGs in black.

Data availability

European Nucleotide Archive: *Lycopodia hypogea*. Accession number [PRJEB72483](https://www.ebi.ac.uk/ena/record/PRJEB72483). The genome sequence is released openly for reuse. The *Lycopodia hypogea* genome sequencing initiative is part of the Aquatic Symbiosis Genomics Project (PRJEB43743) and Sanger Institute Tree of Life Programme (PRJEB43745). All raw sequence data and the assembly have been deposited in INSDC databases. Raw data and assembly accession identifiers are reported in [Table 1](#) and [Table 2](#).

Production code used in genome assembly at the WSI Tree of Life is available at <https://github.com/sanger-tol>. [Table 4](#) lists software versions used in this study.

Author information

Contributors are listed at the following links:

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

Table 4. Software versions and sources.

Software	Version	Source
BEDTools	2.30.0	https://github.com/arq5x/bedtools2
bin3C	0.3.3	https://github.com/cerebis/bin3C
BLAST	2.14.0	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/
BlobToolKit	4.3.9	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.5.0	https://gitlab.com/ezlab/busco
bwa-mem2	2.2.1	https://github.com/bwa-mem2/bwa-mem2
checkM	2015-01-16	https://ecogenomics.github.io/CheckM/
Cooler	0.8.11	https://github.com/open2c/cooler
DIAMOND	2.1.8	https://github.com/bbuchfink/diamond
dRep	3.4.0	https://github.com/MrOlm/drep
fasta_windows	0.2.4	https://github.com/tolkit/fasta_windows
FastK	1.1	https://github.com/thegenemyers/FASTK
Gfastats	1.3.6	https://github.com/vgl-hub/gfastats
GenomeScope2.0	2.0.1	https://github.com/tbenavi1/genomescope2.0
GTDB-Tk	1.2.1	https://github.com/ECogenomics/GTDBTk
Hifiasm	0.19.8-r603	https://github.com/chhylp123/hifiasm
HiGlass	1.13.4	https://github.com/higlass/higlass
MAGScoT	1.0.0	https://github.com/ikmb/MAGScoT
MaxBin	2.2.7	https://sourceforge.net/projects/maxbin/
MerquryFK	1.1.2	https://github.com/thegenemyers/MERQURY.FK
MetaBAT2	2.15-15-gd6ea400	https://bitbucket.org/berkeleylab/metabat
metaMDBG	Pre-release	https://github.com/GaetanBenoitDev/metaMDBG
metaTOR	Pre-release	https://github.com/koszullab/metaTOR
Minimap2	2.24-r1122	https://github.com/lh3/minimap2
MitoHiFi	2	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
PretextSnapshot	0.0.5	https://github.com/sanger-tol/PretextSnapshot
PretextView	1.0.3	https://github.com/sanger-tol/PretextView
Prokka	1.14.5	https://github.com/tseemann/prokka
Seqtk	1.3	https://github.com/lh3/seqtk
Singularity	3.9.0	https://github.com/sylabs/singularity
sanger-tol/ascc	0.1.0	https://github.com/sanger-tol/ascc
sanger-tol/blobtoolkit	0.4.0	https://github.com/sanger-tol/blobtoolkit
sanger-tol/curationpretext	1.4.2	https://github.com/sanger-tol/curationpretext
TreeVal	1.4.0	https://github.com/sanger-tol/treeval
YaHS	1.1a.2	https://github.com/c-zhou/yahs

References

- Altschul SF, Gish W, Miller W, et al.: **Basic Local Alignment Search Tool.** *J Mol Biol.* 1990; **215**(3): 403–410.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Baghdiguian S, Le Goff E, Paradis L, et al.: **Using the carnivorous sponge *Lycopodium hypogaea* as a nonclassical model for understanding apoptosis-mediated shape homeostasis at the organism level.** *Foundations.* 2023; **3**(2): 220–230.
[Publisher Full Text](#)
- Bateman A, Martin MJ, Orchard S, et al.: **UniProt: the Universal Protein Knowledgebase in 2023.** *Nucleic Acids Res.* 2023; **51**(D1): D523–D531.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Benoit G, Raguideau S, James R, et al.: **High-quality metagenome assembly from long accurate reads with metaDBG.** *Nat Biotechnol.* 2024; **42**(9): 1378–1383.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bowers RM, Kyrpides NC, Stepanauskas R, et al.: **Minimum Information about a Single Amplified Genome (MISAG) and a Metagenome-Assembled Genome (MIMAG) of bacteria and archaea.** *Nat Biotechnol.* 2017; **35**(8): 725–731.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at Tree-of-Life scale using DIAMOND.** *Nat Methods.* 2021; **18**(4): 366–368.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Richards E, Rajan J, et al.: **BlobToolKit - interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chaumeil PA, Mussig AJ, Hugenholtz P, et al.: **GTDB-Tk v2: memory friendly classification with the genome taxonomy database.** *Bioinformatics.* 2022; **38**(23): 5315–5316.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Concepcion GT, Feng X, et al.: **Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chevaldonné P, Pérez T, Crouzet JM, et al.: **Unexpected records of 'deep-sea' carnivorous sponges *Asbestopluma hypogaea* in the shallow NE Atlantic shed light on new conservation issues.** *Marine Ecology.* 2015; **36**(3): 475–484.
[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](#)
- Dupont S, Carre-Mlouka A, Domart-Coulon I, et al.: **Exploring cultivable *Bacteria* from the prokaryotic community associated with the carnivorous sponge *Asbestopluma hypogaea*.** *FEMS Microbiol Ecol.* 2014; **88**(1): 160–174.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dupont S, Corre E, Li Y, et al.: **First insights into the microbiome of a carnivorous sponge.** *FEMS Microbiol Ecol.* 2013; **86**(3): 520–531.
[PubMed Abstract](#) | [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.: **Gfstats: conversion, evaluation and manipulation of genome sequences using assembly graphs.** *Bioinformatics.* 2022; **38**(17): 4214–4216.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Godefroy N, Le Goff E, Martinand-Mari C, et al.: **Sponge digestive system diversity and evolution: filter feeding to carnivory.** *Cell Tissue Res.* 2019; **377**(3): 341–351.
[PubMed Abstract](#) | [Publisher 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](#)
- 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](#)
- Huerta-Cepas J, Serra F, Bork P: **ETE 3: reconstruction, analysis, and visualization of phylogenomic data.** *Mol Biol Evol.* 2016; **33**(6): 1635–1638.
[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](#)
- Koltun VM: **Sponge fauna of the north-western Pacific from the shallows to the ultra-abysal depths.** *Proceedings of the Institute of Oceanology, Academy of Sciences of the U.S.S.R.* 1970; **86**: 165–221.
- 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](#)
- Le Goff E, Martinand-Mari C, Belkhir K, et al.: **Molecular complexity and gene expression controlling cell turnover during a digestive cycle of carnivorous sponge *Lycopodium hypogaea*.** *Cell Tissue Res.* 2022; **388**(2): 399–416.
[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, Seppay M, et al.: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Merkel D: **Docker: lightweight Linux containers for consistent development and deployment.** *Linux J.* 2014; **2014**(239): 2.
[Reference Source](#)
- 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.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Snudplot 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.: **Merquary: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schoch CL, Ciufo S, Domrachev M, et al.: **NCBI Taxonomy: a comprehensive update on curation, resources and tools.** *Database (Oxford).* 2020; **2020**: baaa062.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Seemann T: **Prokka: rapid prokaryotic genome annotation.** *Bioinformatics.* 2014; **30**(14): 2068–2069.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Uliano-Silva M, Ferreira JGRN, Krashenninnikova 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](#)
- Vacelet J, Boury-Esnault N: **A new species of carnivorous sponge (Demospongiae: Cladorhizidae) from a Mediterranean cave.** *Bull Inst Royal Sci Nat Belgique Biol.* 1996; **66**(Suppl): 109–115.
[Reference Source](#)
- Vacelet J, Boury-Esnault N, Fiala-Médioni A, et al.: **A methanotrophic carnivorous sponge.** *Nature.* 1995; **377**(6547): 296.
[Publisher Full Text](#)
- Vacelet J, Boury-Esnault N, Le Goff E, et al.: **Spermatogenesis in the carnivorous sponge *Lycopodium hypogaea* (Porifera, Demospongiae).** *Zoomorphology.* 2022; **141**: 1–17.
[Publisher Full Text](#)
- Vacelet J, Fiala-Médioni A, Fisher CR, et al.: **Symbiosis between methane-oxidizing bacteria and a deep-sea carnivorous cladorhizid sponge.** *Mar Ecol Prog Ser.* 1996; **145**(1–3): 77–85.
[Publisher Full Text](#)
- Vasimuddin Md, Misra S, Li H, et al.: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
[Publisher Full Text](#)
- Zhou C, McCarthy SA, Durbin R: **YaHS: Yet another Hi-C Scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:    

Version 1

Reviewer Report 16 March 2026

<https://doi.org/10.21956/wellcomeopenres.28571.r148487>

© 2026 Senevirathna J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jayan Duminda M Senevirathna 

Uva Wellassa University, Badulla, Uva Province, Sri Lanka

The article is interesting, the methodology is clearly explained, and the data appear to be available for reproducibility. However, there are concerns regarding the interpretation of the genome assembly results.

The reported BUSCO completeness is about 68%, which is relatively low for a genome assembly.

The authors should provide a clear explanation for this low level of gene set identification. Possible reasons such as incomplete assembly, fragmented contigs, sequencing limitations, or annotation issues should be discussed to justify this result.

In addition, in Figure 2, the unit of coverage (e.g., sequencing depth, ×) should be clearly indicated on the X-axis to improve the clarity and interpretation of the figure.

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: Aquatic genomics

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

Reviewer Report 14 March 2026

<https://doi.org/10.21956/wellcomeopenres.28571.r148478>

© 2026 Proux-Wéra E. 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.



Estelle Proux-Wéra 

Stockholm University, stockholm, Sweden

In this article the authors present the chromosomal and mitochondrial genomes of the sponge *Lycopodina hypogea*, and studied the metagenome of their specimen.

The methods used for sequencing and assembling are standard and used modern tools. Information on where to find raw data and assemblies is provided.

The metrics of the assembly seem similar to other sponge genomes in term of genome size and metagenome content. The majority of the assembly is in 15 long chromosomal pseudomolecules.

My only concern regarding the assembly metrics is the BUSCO score completeness being a bit low at 69%. Most of the most recent sponge genome assembled by the Wellcome Sanger Institute have a BUSCO score for completeness around 80%. Just a few other sponges have a BUSCO score of 70% or below. Is some part of the assembly missing? Is the BUSCO reference set metazoa_odb10 not great for sponges?

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: Bioinformatics, genome assembly

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

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

Reviewer Report 13 March 2026

<https://doi.org/10.21956/wellcomeopenres.28571.r148482>

© 2026 Hesketh-Best P. 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.



Poppy Hesketh-Best 

IGTP Institut Germans Trias i Pujol, Barcelona, Spain

The authors present in this data note high-quality nuclear and mitochondrial genomes, with the addition of 27 metagenome-assembled genomes (MAGs) bins for *Lycopodina hypogea*, a marine carnivorous sponge.

The study offers a clear rationale and briefly summarises the current understanding of *L. hypogea*, while also providing well-justified and clearly described methodologies. The quality metrics for the sponge genomes were presented and sufficiently detailed.

The authors do not justify or explain why three different specimens were used for the three sequencing strategies (WGS, RNA-seq, and Hi-C). This omission leaves uncertainty regarding potential biological variation among specimens and how such variation might affect the interpretation and integration of the sequencing results. Particularly with the Hi-C+WGS data. Is this due to small sample sizes/DNA availability? I feel this requires some explanation, or demonstration that the variability between the three specimens are negligible.

The for the nuclear and mitochondrial genome data are available and accessible (2026-03-10), and clearly described in the text. The only information that is unfortunately missing is the availability of the MAG genomes that are described. Fasta files for these MAGs could easily be made available in the Zenodo repository, which currently only contains the contamination and completion statistics from checkM.

Overall, the data note provides valuable resources for investigators working on marine eukaryotic genomes/marine sponges, with some utility for marine microbiologists due to the unavailability of the bacterial genomic data.

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?

Yes

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

Partly

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** Genomics, Microbiology**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 12 March 2026

<https://doi.org/10.21956/wellcomeopenres.28571.r148484>

© 2026 Giles E. 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.

**Emily C Giles** 

Cawthron Institute (Ringgold ID: 5732), Nelson, New Zealand

The authors present a chromosome-level nuclear genome, mitochondrial genome, and several meta genome assembled prokaryote assemblies. The subject organism is the iconic carnivorous sponge, *Lycopodina hypogea*, that inhabits the deep sea. This sponge represents a great model system because it is easy to maintain in tanks, unlike many sea sponges.

The authors use different specimens for genome sequencing, Hi-C scaffolding, and RNA sequencing. This is not ideal. It is best when all genomic/transcriptomic data arise from the same individual. The authors could provide justification as to why different individuals were used to generate the data.

The authors give detailed information about the DNA extraction method used to obtain HMW DNA. This is a useful inclusion. While some might see this as trivial detail, obtaining sufficient high quality DNA is often a challenging step when working with marine taxa.

The details of the genome assembly methods are sufficient and easy to read. It is nice to see assembly methods detailed in-text rather than in externally referenced GitHub pages as is the case for the assembly curation step. Evaluation of the assembly curation methods would be improved if more detailed documentation were provided in the rapid-curation link. For example, the ASCC pipeline details give a high level description of the method rationale, steps, and limitations. This is lacking for the rapid-curation methods.

The final assembly length (235 Mb) differs greatly from the Genomescope estimated genome size (700+ Mb). Additionally, completeness of the nuclear assembly is extremely low. The authors could give potential explanations for these findings. Is it possible that true components of the genome were removed during decontamination/ curation, or is it more likely that Genomescope overestimated the genome size? Do the authors recover any of the missing metazoan buscos in

their MAGs or contigs that were removed during curation? How does the assembly length compare to that of other close relatives? Did the authors estimate genome size using flow cytometry?

The assembly scaffolding looks to be of very high quality with the majority of the assembly falls into 15 pseudo chromosomes.

Is there a typo in the line about transcript length? There appears to be an unintended "5".

It is nice to see the taxonomic tree of the metagenomic bins, however the authors could more clearly indicate the quality of the recovered MAGs. I would not consider MAGs with 50% completeness to be "high quality." The coloration of the Figure 7 scale bars are misleading; it is very hard to discern which MAGs have very low (50%) completeness vs. which MAGs have high (>80%) completeness. Furthermore, I only count 24 MAGs in Figure 7. These details need to be clarified in the results and abstract of the manuscript.

Overall, future comparative studies should consider the low completeness of the nuclear and prokaryote genomes presented here, but the findings are valuable resources for the scientific community.

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

Yes

Are the protocols appropriate and is the work technically sound?

Yes

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

Yes

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: comparative genomics, population genomics, speciation, marine science

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.
