Check for updates

DATA NOTE

ERGA-BGE reference genome: *Homalothecium mandonii* - IUCN vulnerable Bryophyte species endemic to Macaronesia

[version 1; peer review: 1 approved]

Manuela Sim-Sim[1,2], Anabela Martins[1,2], Susana Fontinha[1,3], Carlos Lobo[3], Genoscope Sequencing Technical Team, Alice Moussy[4], Corinne Cruaud[4], Karine Labadie[4], Simone Duprat[5], Lola Demirdjian[5], Patrick Wincker[5], Pedro H. Oliveira[5], Jean-Marc Aury[5], Chiara Bortoluzzi[6]

[1]Centre for Ecology, Evolution and Environmental Changes & CHANGE—Global Change and Sustainability Institute, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[2]Museu Nacional de Historia Natural e da Ciência, Lisboa, Portugal
[3]Jardim Botânico da Madeira Engenheiro Rui Vieira, Funchal, Madeira, Portugal
[4]Genoscope, Institut François Jacob, CEA, Université Paris-Saclay, Evry, France
[5]Génomique Métabolique, Genoscope, Institut François Jacob, Commissariat à l'Energie Atomique (CEA), CNRS, Université Evry, Université Paris-Saclay, Evry, France
[6]Swiss Institute of Bioinformatics, Lausanne, Switzerland

v1 **First published:** 10 Sep 2026, **6**:341
https://doi.org/10.12688/openreseurope.24602.1
Latest published: 10 Sep 2026, **6**:341
https://doi.org/10.12688/openreseurope.24602.1

Abstract

The reference genome of *Homalothecium mandonii* provides an important resource for studying the genetic basis of adaptation and conservation of Macaronesian endemic bryophytes. The entirety of the genome sequence was assembled into 10 contiguous chromosomal pseudomolecules, along with one mitochondrial genome and one chloroplast. This chromosome-level assembly encompasses 416.4 Mb, composed of 45 contigs and 20 scaffolds, with contig and scaffold N50 values 23.3 Mb and 38.8 Mb, respectively.

Keywords

Homalothecium mandonii, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Macaronesian Bryophyte

HE This article is included in the Horizon Europe gateway.

Open Peer Review

Approval Status ✓

	1
version 1 10 Sep 2026	✓ view

1. **Hengchi Chen**, Georg-August-University Goettingen, Goettingen, Germany

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

This article is included in the Genome Reports from the Biodiversity Genomics Europe Project collection.

Corresponding authors: Manuela Sim-Sim (mmsim-sim@fc.ul.pt), Anabela Martins (anabelamartins323@gmail.com), Chiara Bortoluzzi (chiara.bortoluzzi@sib.swiss)

Author roles: Sim-Sim M: Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Martins A**: Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Fontinha S**: Investigation, Writing – Review & Editing; **Lobo C**: Investigation, Writing – Review & Editing; **Moussy A**: Investigation, Supervision, Writing – Review & Editing; **Cruaud C**: Investigation, Supervision, Writing – Review & Editing; **Labadie K**: Investigation, Supervision, Writing – Review & Editing; **Duprat S**: Data Curation, Formal Analysis, Writing – Review & Editing; **Demirdjian L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Wincker P**: Investigation, Supervision, Writing – Review & Editing; **H. Oliveira P**: Investigation, Supervision, Writing – Review & Editing; **Aury JM**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Bortoluzzi C**: Formal Analysis, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Biodiversity Genomics Europe (Grant no. 101059492) is funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. This work was supported by the Genoscope, the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), France Génomique (ANR-10-INBS-09-08), and the exploratory research programme 'ATLASea: Atlas of marine genomes and its targeted project SEQ-Sea (ANR-22-EXAT-0003-SEQ-Sea), with additional support from CE3C's FCT Unit funding UID/00329/2025 (https://doi.org/10.54499/UID/00329/2025)."
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Sim-Sim M *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: Sim-Sim M, Martins A, Fontinha S *et al.* **ERGA-BGE reference genome: *Homalothecium mandonii* - IUCN vulnerable Bryophyte species endemic to Macaronesia [version 1; peer review: 1 approved]** Open Research Europe 2026, **6**:341 https://doi.org/10.12688/openreseurope.24602.1

First published: 10 Sep 2026, **6**:341 https://doi.org/10.12688/openreseurope.24602.1

Introduction

Homalothecium mandonii (Mitt.) Geh., a member of the family Brachytheciaceae, is a moss species characterized by distinct morphological traits. This species is endemic to Macaronesia (Azores, Madeira and the Canary Islands), where it grows on rocks, rocky slopes along trails and as an epiphyte on trees and shrubs of natural forests (Hedenäs et al., 2014). Due to habitat degradation, climate change, and the spread of invasive species, *H. mandonii* is currently classified as "Vulnerable" on the IUCN Red List, reflecting the high risk coming from further population declines (Patiño et al., 2016; Hodgetts et al., 2019; Sim-Sim et al., 2019). Beyond its conservation status, *H. mandonii* performs several important ecological functions within natural forests of Macaronesia. As a bryophyte, it contributes to soil stabilization, water retention, microclimate regulation, nutrient cycling, carbon sequestration, and erosion control, while also providing microhabitats for invertebrates such as springtails, mites, and woodlice (Glime, 2024). These ecological roles further emphasize the importance of conserving both the species and its natural habitats. In this context, developing a high-quality reference genome for *H. mandonii* is essential for advancing our understanding of its unique genetic makeup and adaptive traits. Such a genomic resource would support conservation efforts by informing habitat management and restoration strategies, while also providing valuable insights into evolutionary processes and genetic diversity within bryophytes, with broader implications for plant genomic research (Ye et al., 2025).

The generation of this reference resource was coordinated by the European Reference Genome Atlas (ERGA) initiative's Biodiversity Genomics Europe (BGE) project, supporting ERGA's aims of promoting transnational cooperation to promote advances in the application of genomics technologies to protect and restore biodiversity (Mazzoni et al., 2023).

Materials & methods

ERGA's sequencing strategy includes Oxford Nanopore Technology (ONT) and/or Pacific Biosciences (PacBio) for long-read sequencing, along with Hi-C sequencing for chromosomal architecture, Illumina Paired-End (PE) for polishing (i.e. recommended for ONT-only assemblies), and RNA sequencing for transcriptomic profiling, to facilitate genome assembly.

Sample and sampling information

A total of 31 whole gametophyte plants of *Homalothecium mandonii* were sampled from a single population, identified using the key provided by Hedenäs et al. (2014), at the Fanal region, Madeira Island, Portugal (32.808611, −17.141388), on 7 February 2024 by Manuela Sim-Sim, Anabela Martins, Carlos Lobo and Susana Fontinha. Sampling and transport of specimens were authorized under Transport Declaration No. 01/IFCN/2023 and Declaration No. 05/IFCN/2023, issued by Instituto das Florestas e Conservação da Natureza (IFCN), (FLO MAD), as well as Export Declaration 23PTLX01094D, issued by Instituto da Conservação da Natureza e das Florestas (ICNF). The specimens were carefully collected and cleaned using sterilized tweezers and were immediately frozen in liquid nitrogen and stored at −80 °C until RNA extraction.

Vouchering information

Physical reference materials for the sequenced specimen have been deposited in the Herbarium LISU from the Museu Nacional de História Natural e da Ciência, Universidade de Lisboa under the accession number LISU272135.

Frozen reference tissue material of the whole plant from a proxy voucher is available at the Biobank [Bryophyta] [http://biotropical.ihmt.unl.pt] under the voucher ID BTRA1074.

An electronic voucher image of the sequenced individual is available from the BioImageArchive dataset https://www.ebi.ac.uk/biostudies/bioimages/studies/S-BIAD3440 under accession ID S-BIAD3440 and doi:10.6019/S-BIAD3440.

Data availability

Homalothecium mandonii and the related genomic study were assigned to Tree of Life ID (ToLID) 'cbHomMand10', and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB79978. The sample information is available for BioSample accessions SAMEA115291703, SAMEA115291708, and SAMEA115291719. The genome assembly is accessible from ENA under accession number GCA_981437985.1, and the annotated genome will be made available through the Ensembl website when available (https://projects.ensembl.org/erga-bge/). Sequencing data produced as part of this project are available from ENA at the following accessions: ERX15434671, ERX15434349, ERX15435639, ERX15435641, ERX15434351, and ERX15435640. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) document available at https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Homalothecium_mandonii/cbHomMand10. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/378980.

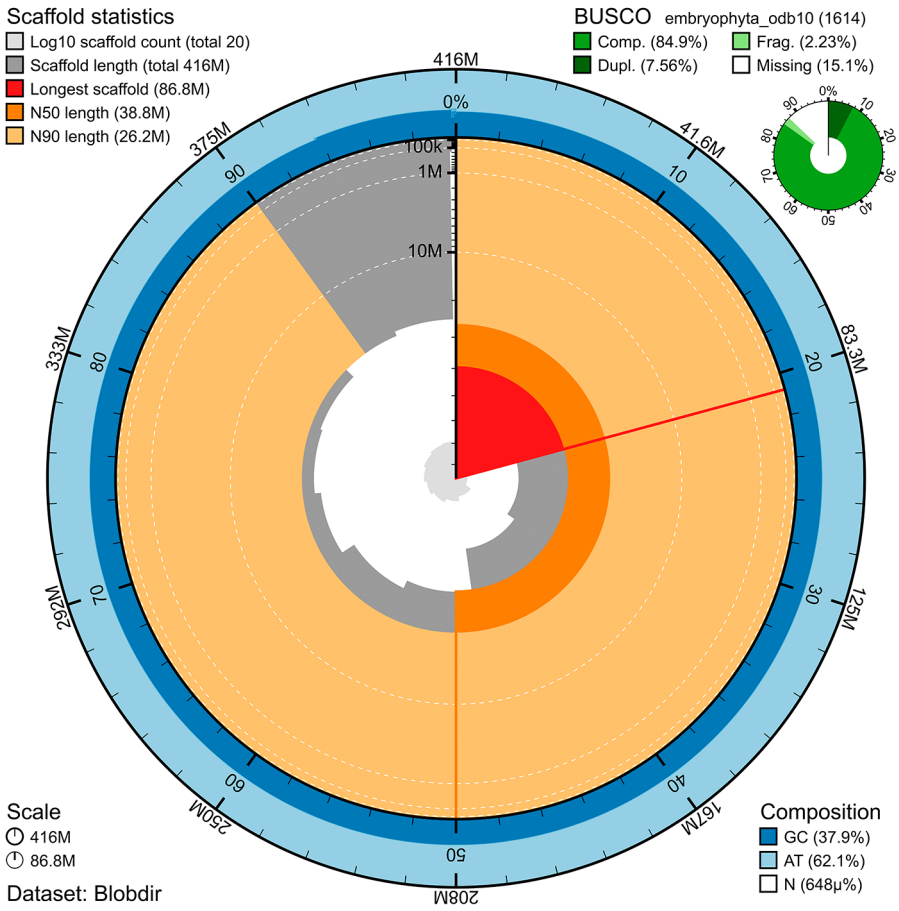


Figure 1. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 416,463,416 bp assembly, including the mitochondrial genome and chloroplast. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (86.8 MB, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (38,796,100 and 26,158,100 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log-scale, with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT, and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated, and missing BUSCO genes found in the assembled genome from the Embryophyta (odb10) database is shown in the top right.

Genetic Information

The estimated genome size, based on ancestral taxa, is 565.1 Mb. This is inferred as a diploid genome with a haploid chromosome number of 10 (2n = 20). The assembled genome spans 416.2 Mb at chromosome level. All information for this species was retrieved from Genomes on a Tree (Challis et al., 2023).

DNA/RNA processing

DNA was extracted from 40 mg of whole plant using the Silex protocol, based on a standard CTAB method with a DNA silica matrix recovery (https://plantmethods.biomedcentral.com/articles/10.1186/s13007-020-00652-y). DNA fragment size selection was performed using Short Read Eliminator (PacBio). Quantification was performed using a Qubit dsDNA HS Assay kit (Thermo Fisher Scientific), and integrity was assessed in a FemtoPulse system (Agilent). DNA was stored at 4 °C until use.

RNA was extracted from a whole plant (8 mg) using RNeasy Powerplant Kit (Qiagen) following manufacturer instructions. Residual genomic DNA was removed with 6 U of TURBO DNase (2 U/μl) (Thermo Fisher Scientific). Quantification was performed using a Qubit RNA HS Assay, and integrity was assessed in a Bioanalyzer system (Agilent). RNA was stored at −80 °C.

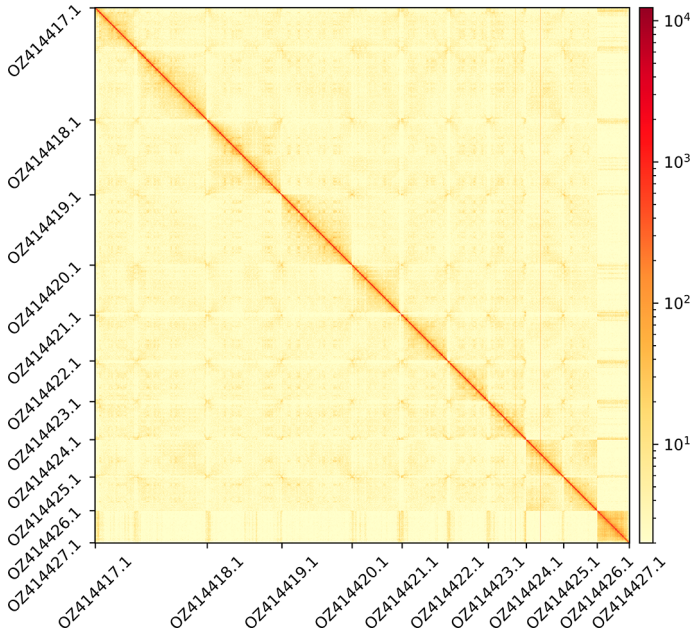


Figure 2. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 200 kb bin size for plotting. For visibility purposes, both axes show the 10 contiguous chromosomal pseudomolecules and the mitochondrial genome (chloroplast genome excluded).

Library preparation and sequencing

Long-read DNA libraries were prepared with SMRTbell prep kit 3.0 following the manufacturer's instructions and sequenced on a Revio system (PacBio). A supplementary purification step using the DNeasy PowerClean Pro Cleanup Kit was performed after DNA fragmentation to optimize sequencing yield.

Hi-C library was generated from another whole plant (10 mg) using the Arima High Coverage HiC Kit (following the Animal Tissues low input protocol v01) and sequenced on a NovaSeq X Plus instrument (Illumina) with 2x 150 read length.

Poly(A) RNA-Seq libraries were constructed using the Illumina Stranded mRNA Prep, Ligation Prep kit (Illumina) and sequenced on an Illumina NovaSeq X Plus instrument.

Genome assembly methods

The genome was assembled using the Genoscope GALOP pipeline (https://workflowhub.eu/workflows/1200). Briefly, raw PacBio HiFi reads were assembled using Hifiasm v0.24.0-r703 (Cheng et al., 2021). Remaining allelic duplications were removed using purge_dups v1.2.5 (Guan et al., 2020) with default parameters and the proposed cutoffs. The purged assembly was scaffolded using YaHS v1.2 (Zhou et al., 2023), and assembled scaffolds were then curated through manual inspection using PretextView v0.2.5 to remove false joins and incorporate sequences not automatically scaffolded into their respective locations within the chromosomal pseudomolecules. Finally, the mitochondrial and chloroplastic genomes were assembled using Oatk (Zhou et al., 2025) and included in the released assembly (GCA_981437985.1). Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (https://doi.org/10.48546/workflowhub.workflow.1104.1).

Results

Genome assembly

The genome assembly has a total length of 416,463,416 bp in 20 scaffolds, including the mitogenome and chloroplast (Figures 1 & 2), with a GC content of 38%. The assembly has a contig N50 of 23,327,363 bp and L50 of 7, and a scaffold N50 of 38,796,100 bp and L50 of 4. The assembly does not have gaps. The single-copy gene content analysis using the Embryophyta (odb10) database with BUSCO v5.8.0 (Manni et al., 2021) resulted in 84.9% completeness (77.3% single

and 7.6% duplicated). 12.69% of reads k-mers were present in the assembly, and the assembly has a base accuracy Quality Value (QV) of 74.26 as calculated by Merqury (Rhie et al., 2020).

Conflict of interest

The authors declare no conflict of interest related to this study. The funding sources had no involvement in the study design, collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. All authors have participated sufficiently in the work to take public responsibility for the content and agree to the submission of this manuscript.

Funder information

This project received funding from Horizon Europe under the Biodiversity, Circular Economy and Environment (REA. B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054; and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme, with additional support from CE3C's FCT Unit funding UID/00329/2025 (https://doi.org/10.54499/UID/00329/2025). This work was supported by the Genoscope, the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), France Génomique (ANR-10-INBS-09-08), and the exploratory research programme 'ATLASea: Atlas of marine genomes and its targeted project SEQ-Sea (ANR-22-EXAT-0003-SEQ-Sea).

Author contributions

Manuela Sim-Sim coordinated the project, Manuela Sim-Sim, Anabela Martins, Susana Fontinha and Carlos Lobo collected and identified the species, Manuela Sim-Sim sampled and preserved biological material and provided metadata, Manuela Sim-Sim and Anabela Martins provided sampling and metadata support and management, Genoscope Sequencing Technical Team extracted DNA, prepared libraries, and performed sequencing under the supervision of Alice Moussy, Corinne Cruaud, Karine Labadie, Pedro H. Oliveira, and Patrick Wincker, Lola Demirdjian and Simone Duprat performed genome assembly and curation under the supervision of Jean-Marc Aury, Chiara Bortoluzzi generated the figures used in the report. All authors contributed to the writing, review, and editing of this genome note and read and approved the final version. This work is part of the species assigned to Genoscope, which was instrumental in the wet lab, sequencing, and assembly processes, and represents a key contribution to BGE's outputs.

Author information

Members of the Genoscope Sequencing Technical Team (GSTT) are listed here: https://doi.org/10.5281/zenodo.14611490.

Acknowledgements

We would like to acknowledge Christian de Guttry for his management support throughout the project. Acknowledgment to the Biobank (Biotropical Resources, BIOTROP, Global Health and Tropical Medicine, Institute of Hygiene and Tropical Medicine, NOVA University Lisbon, GHTM IHMT NOVA) for the collaboration in depositing the samples. We would like to acknowledge the assembly reviewer, Jèssica Gómez-Garrido, from the Centro Nacional de Análisis Genómico (CNAG).

Literature Cited

Challis R, Kumar S, Sotero-Caio C: **Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life [version 1; peer review: 2 approved].** 2023; **8**(24).
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**

Glime JM: **Roles of bryophytes in forest sustainability—Positive or negative?.** *Sustainability.* 2024; **16**: 2359.
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**

Hedenäs L, Désamoré A, Laenen B, *et al.*: **Three species for the price of one within the moss *Homalothecium sericeum* s.l.** *Taxon.* 2014; **63**(2): 249–257.
Publisher Full Text

Hodgetts N, Cálix M, Englefield E, *et al.*: *A miniature world in decline: European Red List of Mosses, Liverworts and Hornworts.* International Union for Conservation of Nature; 2019.
Publisher Full Text

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.** *Mol. Biol. Evol.* 2021; **38**(10): 4647–4654.
Publisher Full Text

Mazzoni C, Ciofi C, Waterhouse R: **Biodiversity: An atlas of European reference genomes.** *Nature.* 2023; **619**: 252–252.
Publisher Full Text

Patiño J, Mateo RG, Zanatta F, *et al.*: **Climate threat on the Macaronesian endemic bryophyte flora.** *Sci. Rep.* 2016; **6**: 29156.
Publisher 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**

Sim-Sim M, Hodgetts N, Martins A, *et al.*: ***Homalothecium mandonii* (Europe assessment).** ***The IUCN Red List of Threatened Species 2019***, e.T85624112A87712596. 2019.
Reference Source | **Publisher Full Text**

Ye H, Luo G, Liu J, *et al.*: **Decoding genetic diversity through genome engineering in bryophytes.** *Plant J.* 2025; **121**: e70103.
PubMed Abstract | **Publisher Full Text**

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
Publisher Full Text

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

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 26 September 2026

https://doi.org/10.21956/openreseurope.26656.r80932

© **2026 Chen H.** 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.

✔ **Hengchi Chen**
Georg-August-University Goettingen, Goettingen, Germany

This manuscript is submitted as Data Note. The listed procedures of specimen preparation, long-read and Hi-C sequencing, and genome assembly are fairly standard. But the authors appear to omit the genome survey analysis which can provide valuable information for the expected genome size.

Is the assembled genome size 416.4 Mb expected? Another question is that the RNA sequencing data did not seem to be used in the genome assembly or there is some Method detail missing.

Other than these, this new genome assembly of *Homalothecium mandonii* is of good quality and valuable in a range of benchmarks.

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

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

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

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Plant Systems Biology, Polyploidy

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.