Check for updates

DATA NOTE

ERGA-BGE reference genome: *Orthotrichum diaphanum* – a globally widespread epiphytic moss common in Europe

[version 1; peer review: 1 approved]

Pablo Aguado-Ramsay[1,2], Francisco Lara[1,2], Christian de Guttry[3], Chiara Bortoluzzi[3], Maria Conejero[4], Astrid Böhne[5], Rita Monteiro[5], Thomas Marcussen[6,7], Torsten Hugo Struck[6], Rebekah A Oomen[6-10], Genomescope Sequencing Team, Alice Moussy[11], Corinne Cruaud[11], Karine Labadie[11], Lola Demirdjian[12], Emilie Téodori[12], Caroline Belser[12], Patrick Wincker[12], Pedro H Oliveira[12], Jean-Marc Aury[12], Isabel Draper[1,2]

[1]Departamento de Biología - Facultad de Ciencias, Autonomous University of Madrid, Madrid, Community of Madrid, ES-28049, Spain
[2]CIBC-UAM Centro de Investigación en Biodiversidad y Cambio Global, Autonomous University of Madrid, Madrid, Community of Madrid, ES-28049, Spain
[3]SIB Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland
[4]Department of Biodiversity and Evolutionary Biology, MNCN-CSIC Museo Nacional de Ciencias Naturales, Madrid, Spain
[5]Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Bonn, 53113, Germany
[6]Natural History Museum, University of Oslo, Oslo, Norway
[7]Centre for Ecological & Evolutionary Synthesis, University of Oslo, Oslo, Norway
[8]Department of Biological Sciences, University of New Brunswick Saint John, Saint John, New Brunswick, Canada
[9]Tjärnö Marine Laboratory, University of Gothenburg, Gothenburg, Sweden
[10]Centre for Coastal Research, University of Agder, Kristiansand, Norway
[11]CEA, CNRS, Univ Evry, Université Paris-Saclay, Institut François Jacob, Genoscope, Evry, 91057, France
[12]Genomique Metabolique, Évry-Courcouronnes, Île-de-France, France

v1 **First published:** 29 Jul 2026, **6**:270
https://doi.org/10.12688/openreseurope.24584.1
Latest published: 29 Jul 2026, **6**:270
https://doi.org/10.12688/openreseurope.24584.1

Abstract

The reference genome of *Orthotrichum diaphanum* Schrad. ex Brid. will provide a foundational genomic resource for phylogenomic, ecological, and evolutionary studies within Orthotrichaceae Arn. and, more broadly, bryophytes. Its ecological amplitude and global distribution makes it especially valuable for comparative genomics and understanding patterns of adaptation across environmental gradients. The genome sequence was assembled into 10 nuclear chromosomal pseudomolecules, plus one mitochondrial and one chloroplast genome. This chromosome-level assembly encompasses 0.30 Gb, composed of 354 contigs and 27 scaffolds, with contig and scaffold N50 values of 1.43 Mb and 29.3 Mb, respectively.

Open Peer Review

Approval Status ✓

	1
version 1 29 Jul 2026	✓ view

1. **Matthew W. Davis**, University of California, Davis, Davis, USA

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

Keywords
Orthotrichum diaphanum, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Bryophyta, moss, Orthotrichaceae

This article is included in the Horizon Europe gateway.

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

Corresponding author: Isabel Draper (isabel.draper@uam.es)

Author roles: Aguado-Ramsay P: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Lara F**: Data Curation, Funding Acquisition, Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **de Guttry C**: Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Bortoluzzi C**: Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Review & Editing; **Conejero M**: Resources, Writing – Review & Editing; **Böhne A**: Resources, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Marcussen T**: Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Struck TH**: Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Oomen RA**: Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Moussy A**: Investigation, Supervision, Writing – Review & Editing; **Cruaud C**: Investigation, Supervision, Writing – Review & Editing; **Labadie K**: Investigation, Supervision, Writing – Review & Editing; **Demirdjian L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Téodori E**: Data Curation, Formal Analysis, Writing – Review & Editing; **Belser C**: Data Curation, Formal Analysis, Writing – Review & Editing; **Wincker P**: Investigation, Supervision, Writing – Review & Editing; **Oliveira PH**: Investigation, Supervision, Writing – Review & Editing; **Aury JM**: Investigation, Supervision, Writing – Review & Editing; **Draper I**: Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, 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. PAR, FL, and ID have received funding by the Spanish Research Agency (grants PID2020-115149GB-C21, PID2020-115149GB-C22, PID2024-161184NB-C21 and PID2024-161184NB-C22). PAR received financial support from the Autonomous Community of Madrid (PIPF-2022/ECO-24575). 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).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Aguado-Ramsay P *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Aguado-Ramsay P, Lara F, de Guttry C *et al.* **ERGA-BGE reference genome: *Orthotrichum diaphanum* – a globally widespread epiphytic moss common in Europe [version 1; peer review: 1 approved]** Open Research Europe 2026, **6**:270 https://doi.org/10.12688/openreseurope.24584.1

First published: 29 Jul 2026, **6**:270 https://doi.org/10.12688/openreseurope.24584.1

Introduction

Orthotrichum diaphanum is a small epiphytic moss belonging to the family Orthotrichaceae. The species was described by Bridel (1801), shortly after the establishment of bryophyte taxonomy as a formal discipline following the work of Hedwig (1801). Due to its extraordinary ecological amplitude and wide distribution, *O. diaphanum* has historically been described under numerous infra-specific taxa (Brinda and Atwood, 2026), including varieties and forms based on morphological and ecological variation. Most of these names are currently regarded as taxonomic synonyms, with only two varieties generally accepted in recent treatments (Lewinsky, 1993; Lara et al., 2016), the type variety and *O. diaphanum* var. *podocarpi* (Müll. Hal.) Lewinsky from South America.

The species is easily distinguished from other members of Orthotrichaceae by the presence of a conspicuous, long, hyaline leaf acumen (Figure 1), which provides a clear and reliable diagnostic character even in the field. Additional distinguishing traits include its small size, the formation of often extensive dark olive-green to nearly blackish turfs, and a sporophyte bearing a peristome with 16 white free exostome teeth, and 16 endostome segments. The species is usually corticolous, growing mainly on trunks and branches of a wide variety of trees and shrubs, but it also frequently colonizes artificial substrates such as concrete, walls, and other urban structures. This ecological plasticity and its tolerance to dryness and nitrified substrates makes *O. diaphanum* particularly abundant in urban and suburban environments, where it is often one of the dominant epiphytic mosses.

Orthotrichum diaphanum has a wide intercontinental distribution. It occurs throughout Europe and western Asia, where it is among the most common epiphytic mosses, being present in almost all countries and major islands. Beyond that, the species is widely distributed across North and South America and is also recorded from South Africa. Its broad disjunct distribution suggests a strong capacity for long-distance spore dispersal and successful establishment under a wide range of climatic conditions. Its ecological plasticity and common occurrence make it a pivotal species in studies of epiphytic moss ecology. At the European level, *O. diaphanum* is classified as Least Concern (LC; Hodgetts and Lockhart, 2020), reflecting its wide distribution and abundance.

This species is autoicous. Karyotype analyses within *Orthotrichum* Hedw. usually reveal a haploid chromosome number of n = 10, including records attributed to *O. diaphanum* (Rice et al., 2015).

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 and annotation.

Sample and sampling information

One specimen of *Orthotrichum diaphanum* was sampled and species identity was determined based on morphology using Lara & Garilleti (2014). The biological material collected in Spain, and used to generate digital sequences, was retrieved from wildlife taxa regulated by the Spanish Royal Decree 124/2017 (https://www.boe.es/eli/es/rd/2017/02/24/124). No ABS permits were granted as utilization of the material falls under exemption in Article 3(2) of the Spanish Royal Decree 124/2017 and was confirmed by the national authorities. The individual was harvested from the tree and was taken directly to the laboratory. Clean apical parts of gametophores, whose taxonomic identity had been double-checked, were individually selected, avoiding sporophytes and reducing contamination risk. Shoots were then weighed and preserved at −80 °C until DNA extraction.

Vouchering information

The colony was deposited in the herbarium of the Universidad Autónoma de Madrid https://sweetgum.nybg.org/science/ih/herbarium-details/?irn=172673 under voucher ID MAUAM-Brio 5488.

Frozen reference tissue material is available from the same individual at the Museo Nacional de Ciencias Naturales https://www.mncn.csic.es/es under voucher ID MNCN-ADN 151725.



Figure 1. ***Orthotrichum diaphanum*** **(MAUAM-Brio 5488).**

Genetic information

The estimated genome size, based on ancestral taxa, is 0.34 Gb. This corresponds to a haploid genome with a haploid number of 10–11 chromosomes. The genome that was sequenced here is a haploid genome with n = 10 and a genome size of 0.30 Gb, which aligns with the estimated values. All information for this species was retrieved from Genomes on a Tree (Challis et al., 2023).

DNA/RNA processing

DNA was extracted from approximately 200 mg of modular colony tissue ground in liquid nitrogen with mortar and pestle. The resulting powder was incubated in 10 mL of extraction buffer (100 mM Tris–HCl, 20 mM EDTA, 1.5 M NaCl, 1% sarkosyl, 2% polyvinylpyrrolidone [PVP], and 0.2% β-mercaptoethanol) at 21 °C for 1 h under agitation (300 rpm). DNA purification was subsequently performed following the general procedure described by Ramakrishnan et al. (2017). DNA fragment size selection was performed using Short Read Eliminator (PacBio). DNA quantification was performed using a Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific), and DNA integrity was assessed using a Genomic DNA 165 Kb Kit (Agilent) on the Femto Pulse system (Agilent). The DNA was stored at +4 °C until used.

Table 1. Chromosomal pseudomolecules in the genome assembly of *Orthotrichum diaphanum*.

NCBI Accession	Chromosome	Size (Mb)
OZ370044.1	1	47.6
OZ370045.1	2	41.7
OZ370046.1	3	41.8
OZ370047.1	4	29.3
OZ370048.1	5	24.7
OZ370049.1	6	23.7
OZ370050.1	7	24.0
OZ370051.1	8	22.9
OZ370052.1	9	21.6
OZ370053.1	10	21.7
OZ370054.1	MIT	0.105
OZ370055.1	PLTD	0.123

RNA was extracted from a modular colony (50 mg) using RNeasy PowerPlant Kit (Qiagen) following the 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 until used.

Library preparation and sequencing

The Long-read DNA library was prepared with SMRTbell prep kit 3.0 following the manufacturers' instructions and sequenced on a Revio system (PacBio). Hi-C library was generated from the same colony used for DNA extraction using the Arima High Coverage HiC Kit (following the Animal Tissues low input protocol v01) and sequenced on NovaSeq X Plus instrument (Illumina) with 2x 150 bp length. In total 39x HiFi and 150x HiC data were sequenced to generate the assembly.

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 NextDenovo v2.5.2 (Hu et al., 2024). Remaining allelic duplications were removed using purge_dups v1.2.5 (Guan et al., 2020) with default parameters and the proposed cutoffs. This assembly was scaffolded using YaHS v1.2.2 (Zhou et al., 2023) and assembled scaffolds were then curated through manual

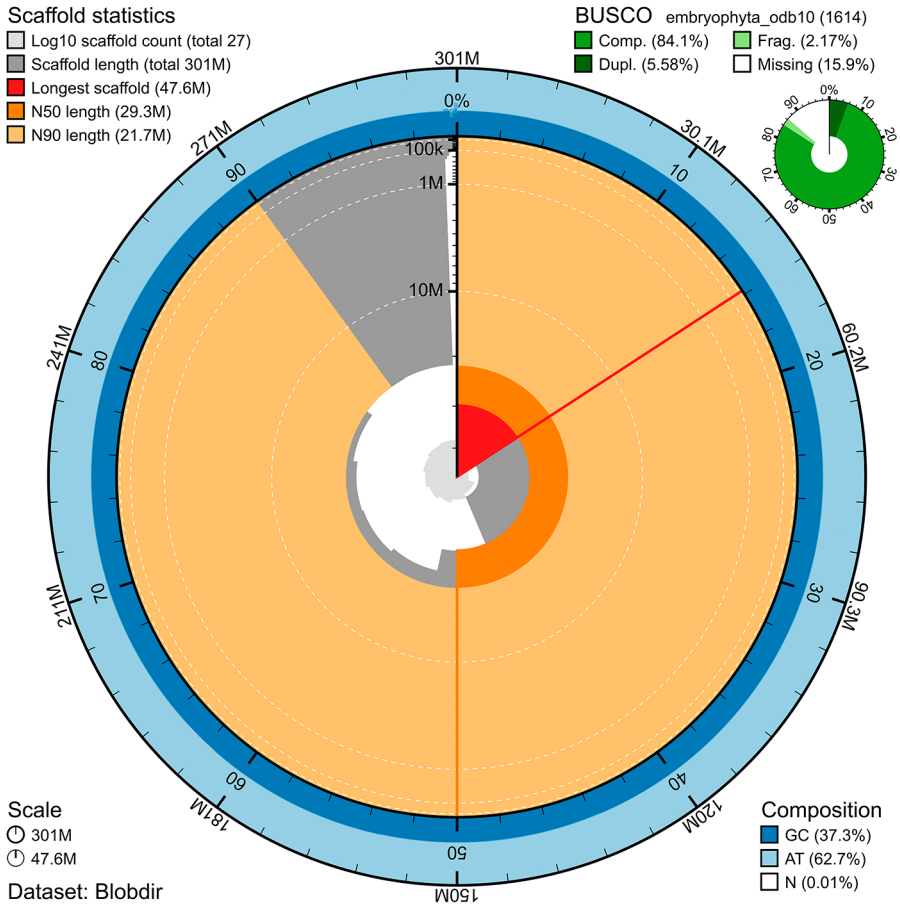


Figure 2. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 300,855,570 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (47.6 Mb, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (29,290,605 and 21,724,835 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 database (odb10) is shown in the top right.

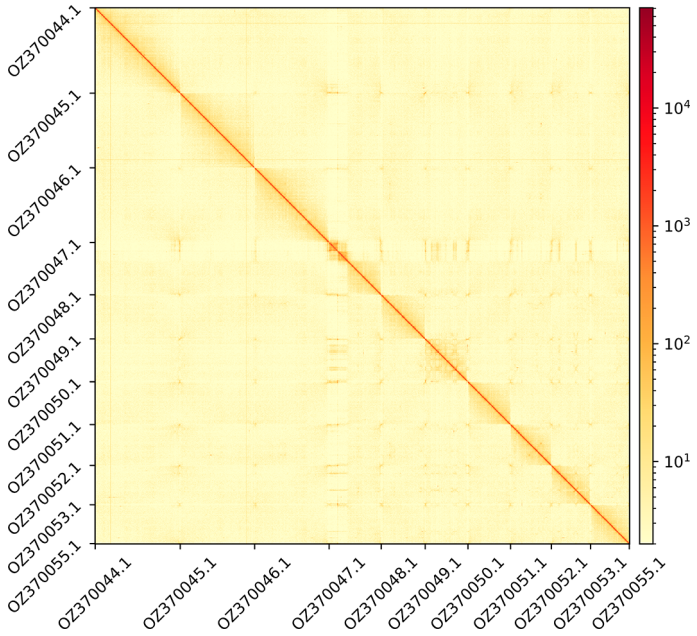


Figure 3. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 100 kb bin size for plotting.

inspection using PretextView v0.2.5 (Harry, 2022) to remove false joins and incorporate sequences not automatically scaffolded into their respective locations within the chromosomal pseudomolecules. Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. Finally, the mitochondrial and chloroplastic genomes were assembled respectively as two single contigs of 104,738 bp and 123,118 bp using the Oatk v1.0 (Zhou et al., 2025), and are included in the released assembly (GCA_977066345.2) (Table 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 300,855,570 bp in 27 scaffolds including the mitogenome and chloroplast (Figures 2 and 3), with a GC content of 37.5%. The assembly has a contig N50 of 1,432,891 bp and L50 of 60 and a scaffold N50 of 29,290,605 bp and L50 of 4. The assembly has a total of 327 gaps, totalling 38.7 kb in cumulative size. The single-copy gene content analysis using the Embryophyta database (odb10) with BUSCO v5.8.0 (Manni et al., 2021) resulted in 84.1% completeness (78.6% single-copy and 5.6% duplicated). The relatively high proportion of missing BUSCO genes (13.4%) reflects a bias in the Embryophyta database towards vascular plants; such values are therefore common when bryophytes are assessed using this lineage. 87.1% of reads *k*-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 45.3 as calculated by Merqury (Rhie et al., 2020).

Author contributions

PAR coordinated the project; PAR and FL collected and identified the species; PAR sampled and preserved biological material and provided metadata; AsB, CdG, RM, TM, THS, RAO, PAR, ID, and MC provided support in sampling, shipping of biological material, metadata collection, and management; GST extracted DNA, prepared libraries, and performed sequencing under the supervision of AM, CC, KL, PW, and PHO; CBe, LD and ET performed genome assembly and curation under the supervision of JMA; CBo generated the analysis and 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 Team are listed here: https://zenodo.org/records/14611490.

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

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. PAR, FL, and ID have received funding by the Spanish Research Agency (grants PID2020-115149GB-C21, PID2020-115149GB-C22, PID2024-161184NB-C21 and PID2024-161184NB-C22). PAR received financial support from the Autonomous Community of Madrid (PIPF-2022/ECO-24575). 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). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Data availability

Orthotrichum diaphanum and the related genomic study were assigned to Tree of Life ID (ToLID) 'cbOrtDiap4', and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB79987. The sample information is available at the following ENA accessions: SAMEA115283924 and SAMEA119567897. The genome assembly is accessible from ENA under accession number GCA_977066345.2 and the annotated genome will be available through the Ensembl website (https://projects.ensembl.org/erga-bge/). Sequencing data produced as part of this project are available from ENA at the following accessions: ERX15434347, ERX15150794, ERX15434348, and ERX15434350. 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/Orthotrichum_diaphanum/cbOrtDiap4. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/522324.

Acknowledgements

We acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de. NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de. NBI Freiburg. We would like to acknowledge the assembly reviewer, Jo Collins from the Wellcome Sanger Institute.

Literature cited

Bridel SÉ v: *Muscologia Recentiorum.* Gotha/Paris: C. G. Ettinger/Barrois Fils; 1801; **2**(2): 192pp. .

Brinda JC, Atwood JJ, editors. **A synopsis of *Orthotrichum.*** *The Bryophyte Nomenclator.* 2026. [Accessed: 19 Jan 2026].

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

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

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

Hedwig J: **Species muscorum frondosorum: descriptae et tabulis aeneis LXXVII coloratis illustratae.** *Opus posthumum. Sumtu Joannis Ambrosii Barthii, Lipsiae and apud Amand Koenig, Parisiis.* 1801; 353 pp. (+ vi + lxxvii Tab.).

Hodgetts N, Lockhart N: *Checklist and country status of European bryophytes–update 2020. Irish Wildlife Manuals, No. 123.* National Parks and Wildlife Service, Department of Culture, Heritage and the Gaeltacht, Ireland; 2020.

Hu J, Wang Z, Sun Z, *et al.*: **NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads.** *Genome Biol.* 2024; **25**(1): 107.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Lara F, Garilleti R, Goffinet B, *et al.*: ***Lewinskya,* a new genus to accommodate the phaneroporous and monoicous taxa of *Orthotrichum* (Bryophyta, Orthotrichaceae).** *Cryptogamy Bryol.* 2016; **37**(4): 361–382.
Publisher Full Text

Lara F, Garilleti R: **Orthotrichum.** Guerra J, Cano MJ, Brugués M, editors. *Flora Briofítica Ibérica. Vol. V: Orthotrichales: Orthotrichaceae; Hedwigiales: Hedwigiaceae; Leucodontales: Fontinalaceae, Climaciaceae, Anomodontaceae, Cryphaeaceae, Leptodontaceae, Leucodontaceae, Neckeraceae; Hookeriales: Hypopterygiaceae, Hookeriaceae, Leucomiaceae, Pilotrichaceae.* Murcia: UMU-SEB; 2014; 50–135.

Lewinsky J: **A synopsis of the genus *Orthotrichum* Hedw. (Musci, Orthotrichaceae).** *Bryobrothera.* 1993; **2**: 1–59.

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

Mazzoni CJ, Ciofi C, Waterhouse RM: **Biodiversity: an atlas of European reference genomes.** *Nature.* 2023; **619**(7969): 252.
Publisher Full Text

Ramakrishnan GS, Fathima AA, Ramya M: **A rapid and efficient DNA extraction method suitable for marine macroalgae.** *3 Biotech.* 2017; **7**(6): 364.
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**

Rice A, Glick L, Abadi S, *et al.*: **The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers.** *New Phytol.* 2015; **206**(1): 19–26.
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.
Publisher Full Text

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

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 04 September 2026

https://doi.org/10.21956/openreseurope.26636.r79461

© **2026 Davis M.** 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.

✔ **Matthew W. Davis**
University of California, Davis, Davis, USA

Article summary:
The article describes a genome assembly of *Orthotrichum diaphanum*, a common bryophyte. The authors use current best practices for genome assembly and cutting edge data generation techniques. They also deposited a voucher to the herbarium, an important and often overlooked component of archiving plant biology. Overall, the assembled genome is of good quality and there are a few minor comments to address.

Reviewer comments:
The authors discuss RNA sequencing, however there is nothing presented in the manuscript on this data. The authors mention in Data Availability that the annotated genome will be available through the Ensembl website, so I assume the results of the RNA sequencing will be available in the future.

The Hi-C contact map has some strange patterns that may be worth investigating. The end of pseudo-molecule Oz370047.1 has high contact with itself and lower background contact than other sequences. It looks like it may need to be split and associated with both Oz370046.1 and Oz370047.1, though I am not positive. Additionally, Oz370049.1 has an interesting pattern of hot and cold contact spots with itself. Finally, the authors discuss the assembly as having 10 pseudo chromosomes. The contact map shows 10 chromosomes, but there are 11 listed sequences on the axes. The 11th is the plastid genome and does not appear to be represented in the contact map.

Lines with comments and questions:
"...*O. diaphanum* var. *podocarpi* (Müll. Hal.) Lewinsky from South America."
- Is the "Lewinsky" a misformatted citation in the introduction?

Questions for authors:
Will the RNA data be available with a later publication?

According to the chromosome table, the 11th sequence in the contact map is the plastid genome. It does not look like it is a part of the figure, just somehow ended up listed on the axis. Should you remove this?

Can you explain a bit more what may be happening with the odd contact pattern in Oz370047.1? It looks like it's possibly misassembled and should be split to Oz370046.1 and Oz370047.1.

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: Plant biology, 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.