Check for updates

DATA NOTE

ERGA-BGE reference genome of the bladderpod *Alyssoides utriculata* (L.) Medik. (Brassicaceae), a yellow-flowering shrub native to Switzerland

[version 1; peer review: awaiting peer review]

Christian de Guttry[1], Adrien Favre[2], Astrid Böhne[3], Rita Monteiro[3], Thomas Marcussen[4,5], Rebekah A. Oomen[4-8], Torsten H. Struck[5], Chiara Natali[9], Chiara Vergata[9], Maria Angela Diroma[9], Claudio Ciofi[9], Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory Team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Caroline Howard[10], Mark Blaxter[10], Marcella Sozzoni[9], Tom Brown[11,12], Robert M. Waterhouse[1], Chiara Bortoluzzi[1]

[1]SIB Swiss Institute of Bioinformatics, Amphipôle, Quartier UNIL-Sorge, Lausanne, 1015, Switzerland
[2]Parc naturel régional de la Vallée du Trient, Rue Marconi 1, Salvan, 1922, Switzerland
[3]Museum Koenig Bonn, Leibniz Institute for the Analysis of Biodiversity Change, Bonn, 53113, Germany
[4]Centre for Ecological & Evolutionary Synthesis, University of Oslo, Oslo, Norway
[5]Natural History Museum, P.O. Box 1172, Blindern, University of Oslo, Oslo, 0318, Norway
[6]Tjärnö Marine Laboratory, University of Gothenburg, Gothenburg, Sweden
[7]Centre for Coastal Research, University of Agder, Kristiansand, Norway
[8]Department of Biological Sciences, University of New Brunswick Saint John, Saint John, Canada
[9]Department of Biology, Sesto Fiorentino (FI), University of Florence, Florence, 50019, Italy
[10]Tree of Life, Wellcome Sanger Institute, Hinxton, CB10 1SA, UK
[11]Leibniz Institute for Zoo- and Wildlife Research, Berlin, Germany
[12]Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Berlin, Germany

v1 **First published:** 15 Oct 2026, **6**:425
https://doi.org/10.12688/openreseurope.25414.1
Latest published: 15 Oct 2026, **6**:425
https://doi.org/10.12688/openreseurope.25414.1

Abstract

We present the chromosome-level genome assembly for bladderrpod, *Alyssoides utriculata* (Brassicales: Brassicaceae), the only currently accepted species in the genus *Alyssoides.* The phased assembly comprises two haplotypes with total lengths of 1.01 Gb and 0.98 Gb, respectively. For each haplotype, eight chromosome-scale pseudomolecules were recovered. Haplotype 1 comprises 310 contigs assembled into 228 scaffolds, with contig and scaffold N50 values of 108.1 Mb and 135.9 Mb, respectively. Haplotype 2 comprises 181

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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

contigs assembled into 96 scaffolds, with contig and scaffold N50 values of 73.2 Mb and 131.5 Mb, respectively. The assembly also includes the chloroplast genome. This haplotype-resolved reference genome provides a genomic resource for A. utriculata and a foundation for future genomic and evolutionary studies of the species.

Keywords
Alyssoide utriculata, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth BioGenome Project, Alysson renflé

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 authors: Christian de Guttry (christian.deguttry@sib.swiss), Chiara Bortoluzzi (chiara.bortoluzzi@sib.swiss)

Author roles: de Guttry C: Investigation, Project Administration, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Favre A**: Investigation, Resources, Writing – Review & Editing; **Böhne A**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Marcussen T**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Oomen RA**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Struck TH**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Natali C**: Investigation, Writing – Review & Editing; **Vergata C**: Investigation, Writing – Review & Editing; **Diroma MA**: Data Curation, Formal Analysis, Writing – Review & Editing; **Ciofi C**: Investigation, Supervision, Writing – Review & Editing; **Howard C**: Investigation, Supervision, Writing – Review & Editing; **Blaxter M**: Investigation, Supervision, Writing – Review & Editing; **Sozzoni M**: Data Curation, Formal Analysis, Writing – Review & Editing; **Brown T**: Data Curation, Formal Analysis, Writing – Review & Editing; **Waterhouse RM**: Project Administration, Writing – Review & Editing; **Bortoluzzi C**: Visualization, 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.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 de Guttry C *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: de Guttry C, Favre A, Böhne A *et al.* **ERGA-BGE reference genome of the bladderpod *Alyssoides utriculata* (L.) Medik. (Brassicaceae), a yellow-flowering shrub native to Switzerland [version 1; peer review: awaiting peer review]** Open Research Europe 2026, **6**:425 https://doi.org/10.12688/openreseurope.25414.1

First published: 15 Oct 2026, **6**:425 https://doi.org/10.12688/openreseurope.25414.1

Introduction

Alyssoides utriculata (L.) Medik., known as bladderpod or inflated bladderseed, is a perennial flowering plant in the family Brassicaceae (Brassicales) and the only currently accepted species of the genus *Alyssoides* (Figure 1) (Plants of the World Online, 2026). Its native range is heavily fragmented, extending from southern and southeastern Europe to the Caucasus and Türkiye, including relatively isolated populations in the western Alps (Plants of the World Online, 2026). The species typically occurs in dry, rocky habitats, where plants reach approximately 20–50 cm in height. It produces yellow flowers and is also cultivated as an ornamental plant. Cytological studies consistently report a diploid chromosome number of 2n = 16 (Marhold 2006). In Switzerland, *A. utriculata* occurs naturally mainly in the lower Rhône Valley and its lateral valleys in Valais, with occurrences elsewhere in the country largely associated with cultivation (InfoFlora, 2026). A large proportion of the species' stock in this area, where the specimen used for genome sequencing in this study was collected, is on the territory of the Parc Naturel Régional de la Vallée du Trient, which therefore bears a strong responsibility for its conservation. The species inhabits rock crevices, scree, and dry grasslands and is currently classified as Near Threatened (NT) on the Swiss national Red List (InfoFlora, 2026). Here, we provide the first reference genome for *A. utriculata.* The generation of this genomic resource was coordinated through the Biodiversity Genomics Europe (BGE) project of the European Reference Genome Atlas (ERGA) initiative. Through coordinated sampling, sequencing, and genome assembly across European biodiversity, ERGA seeks to expand the availability of high-quality reference genomes and facilitate their use in biodiversity research and conservation (Mazzoni et al., 2023).

Materials & methods

Sample and sampling information

On 11 May 2023, Christian de Guttry collected two adult specimens of *Alyssoides utriculata* in Vernayaz, Canton of Valais, Switzerland. The specimens were identified morphologically by Adrien Favre from the Parc Naturel Régional de la Vallée du Trient. Leaf, flower, and fruit tissues were dissected using a scalpel or scissors, transferred immediately to cryotubes, and preserved in liquid nitrogen until DNA extraction. Sampling was conducted under permit ID273, issued by the cantonal authorities of the Canton of Valais, Switzerland.

Vouchering information

Physical reference materials for the sequenced specimen have been deposited in the Musée de la nature du Valais (https://musee-nature-valais.ch/) under accession number SION008245. No Frozen reference material is available.

Data availability

The sampled specimen of *A. utriculata* and the related genomic study were assigned Tree of Life ID (ToLID) 'ddAlyUtri1' and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB77645. The sample information is available at the following BioSample accessions: SAMEA114530580. Haplotype 1 is accessible from ENA under accession number GCA_977000315.1, while haplotype 2 is available under accession number GCA_977000305.1. The annotated genome will be made available through Ensembl (https://projects.ensembl.org/erga-bge/). Sequencing data produced as part of this project are available from ENA at the following accessions: ERX12722006, ERX13270901, and ERX14241336. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report document available at https://github.com/ERGA-consortium/EARs/blob/main/Assembly_Reports/Alyssoides_utriculata/ddAlyUtri1/. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/457762.

DNA and RNA isolation

Approximately 100 mg of leaves were ground in a mortar with liquid nitrogen. DNA was extracted using the Gemomic TIP (100/G) & Genomic DNA Buffer Set (Qiagen) following the manufacturer's protocol. Quantification was performed using a Qubit 1X dsDNA HS kit (Thermo Fisher Scientific), and integrity was assessed in a Pippin Pulse pulsed-field gel electrophoresis system (Sage Science). DNA was stored at -80 °C until use. Approximately 100 mg of leaves were ground in a mortar with liquid nitrogen, followed by RNA extraction using the Quick-RNA MicroPrep Kit (Zymo Research) using the manufacturer's protocol. Quantification was performed using a Qubit RNA BR assay kit (Thermo Fisher Scientific). RNA integrity was assessed by microfluidics-based chip electrophoresis using an RNA 6000 Nano Kit in a 2100 Bioanalyzer system (Agilent Technologies). RNA was preserved at -80°C until library preparation.

Long-read library preparation and sequencing

Whole DNA was sheared in a Megaruptor 2 DNA shearing system (Diagenode) using large fragment hydropores with a target mean fragment length of 20 kb. Fragment profile was assessed by capillary electrophoresis on a 5200 Fragment Analyser system using an HS large fragment 50 kb kit (Agilent Technologies). Repair and A-tailing of DNA fragments and SMRTbell adapters ligation were performed according to the SMRTbell prep kit 3.0 protocol (Pacific Biosciences). A BluePippin automated pulsed-field gel electrophoresis system (Sage Science) was employed to remove fragments



Figure 1. A specimen of the species *Alyssoides utriculata*. The individual in the picture is from Vernayaz, Canton of Valais, Switzerland, and is the one sampled and sequenced for this study. Photo credit: Christian de Guttry.

shorter than 10 kb. Primer annealing, polymerase binding and preparation of internal DNA control were performed using the Pacific Biosciences Sequel II binding kit and DNA internal control complex 3.2 according to the manufacturer's protocol. The library was sequenced in HiFi mode in a Pacific Biosciences Sequel IIe platform using a Sequel II

Table 1. Genome assembly metrics.

Genome assembly	Haplotype 1	Haplotype 2
Assembly name	ddAlyUtri1.hap1	ddAlyUtri1.hap2
Assembly accession	GCA_977000315.1	GCA_977000305.1
Assembly level	Chromosome	Chromosome
Span (bp)	1,012,219,214	987,112,815
Number of chromosomes	8	8
Number of contigs	310	181
Contig N50 (bp)	108,093,000	73,274,071
Number of scaffolds	228	96
Scaffold N50 (bp)	135,936,682	131,581,310
% Assigned to chromosomes	98.7%	99.3%
Chloroplast genome (bp)	153,111	-

Sequencing plate 2.0 and two 8M ZMW SMRT cells with a 30-hour movie time and 2 hours of pre-extension time for a target 25X genome coverage.

Transcriptome sequencing

RNA Library preparation was carried out using the Illumina Stranded mRNA library prep kit, starting from 700 ηg of RNA and following the manufacturer's instructions. Whole RNA was incubated for messenger RNA (mRNA) capture using poly-T probes. Messenger RNA was purified, fragmented and converted into cDNA, keeping the strand information (+/-). The library was then indexed using a combination of unique dual indexes. The library was quantified using a Qubit RNA HS assay kit (Thermo Fisher Scientific), and quality was checked on a 5200 Fragment Analyser System. A normal distribution with peaks between 300 and 400 bp was expected. The library was pooled in an equimolar

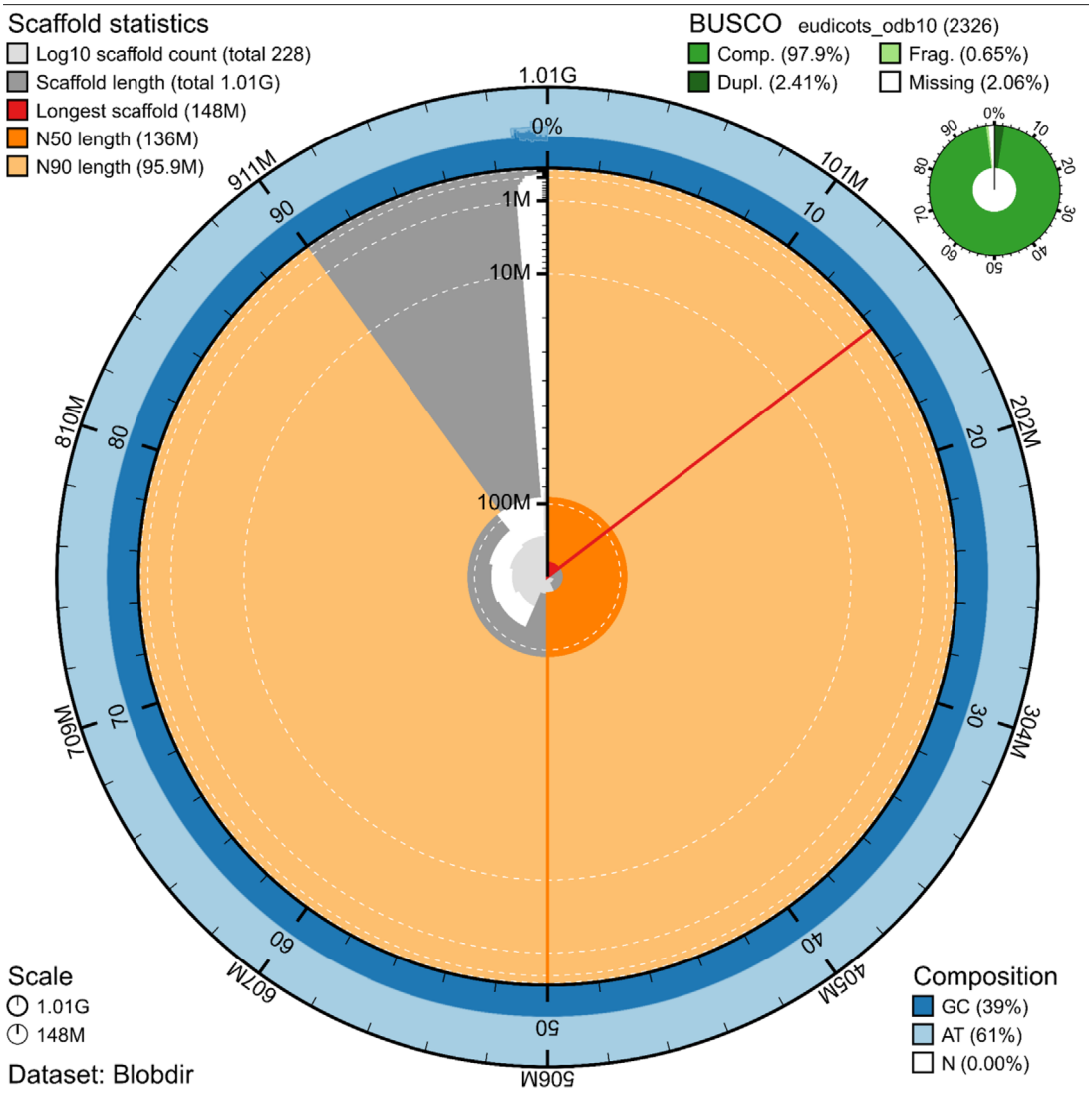


Figure 2. Snail plot summary of assembly metrics for ddAlyUtri1.hap1. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 1,012,219,214 bp assembly including the plastid genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (148 Mb, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (135,936,682 and 95,879,528 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 Eudicots database (odb10) is shown in the top right. The figure was generated using Blobtoolkit (Challis et al., 2020).

ratio with other cDNA libraries and sequenced on an Illumina Novaseq 6000 platform using a NovaSeq 6000 Reagent Kit v1.5 (300 cycles).

Hi-C sequencing

The Hi-C sample was prepared from 20–50 mg of frozen tissue from the ddAlyUtri1 sample, using Qproteome nuclei isolation (Qiagen) followed by the Arima-HiC v2 kit (Arima Genomics). The tissue was homogenised via cryogenic bead beading using the FastPrep, and nuclei isolation was performed. Following the manufacturer's instructions, tissue was fixed, and DNA crosslinked using TC buffer to a final formaldehyde concentration of 2%. 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. 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. The library was sequenced on an Illumina Novaseq instrument.

Genome assembly methods

In total, 32x HiFi and 130x Hi-C were sequenced to generate the assembly. The genome was assembled using the ERGA HiFi & HiC assembly pipeline (De Panis 2024). Briefly, Illumina reads were preprocessed for quality and length using fastp v0.23.4 (Chen et al., 2018), and PacBio HiFi reads were preprocessed for adapter sequences using cutadapt v4.9 (Martin 2011) and initial contigs were assembled using HiFiasm v0.19.9 (Cheng et al., 2021), followed by removal of retained haplotigs using purge-dups v1.2.6 (Guan et al., 2020) and scaffolding with YaHS v1.2a.2 (Zhou et al., 2023). Contaminant sequences were then identified and removed using fcs-gx v0.5.0 (Tvedte et al., 2024). Finally, assembled scaffolds were curated via manual inspection using Pretext v0.2.5 (Harry 2022) with the Rapid Curation Toolkit (https://gitlab.com/wtsi-grit/rapid-curation) to remove any false joins and incorporate any sequences not automatically scaffolded into their respective locations in the chromosomal pseudomolecules (or super-scaffolds). The chloroplast genome was assembled as a single contig of 153,111 bp using oatk v1.0 (Zhou et al. 2025) and included in the released assembly (GCA_977000315.1). We used both oatk and mitoHiFi (Uliano-Silva et al., 2023) to assemble the mitochondrial genome using the HiFi reads, but we were unable to recover the full mitochondrial genome.

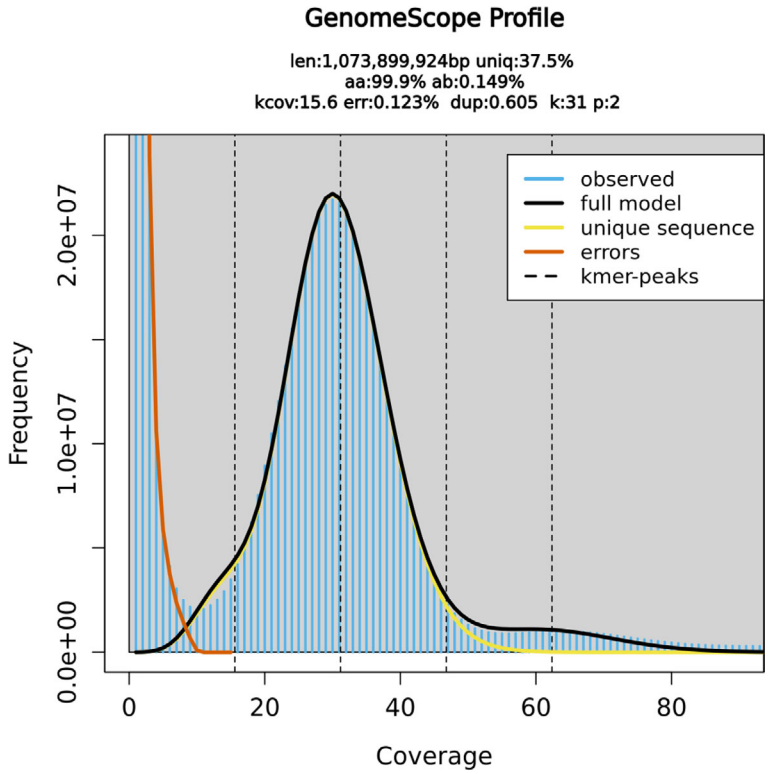


Figure 3. Frequency distribution of *k*-mers. The plot shows observed and modelled *k*-mer spectra of HiFi sequencing data, providing estimates of genome size, heterozygosity, and repeat content. The frequency distribution was generated using GenoScope2 (Ranallo-Benavidez et al., 2020).

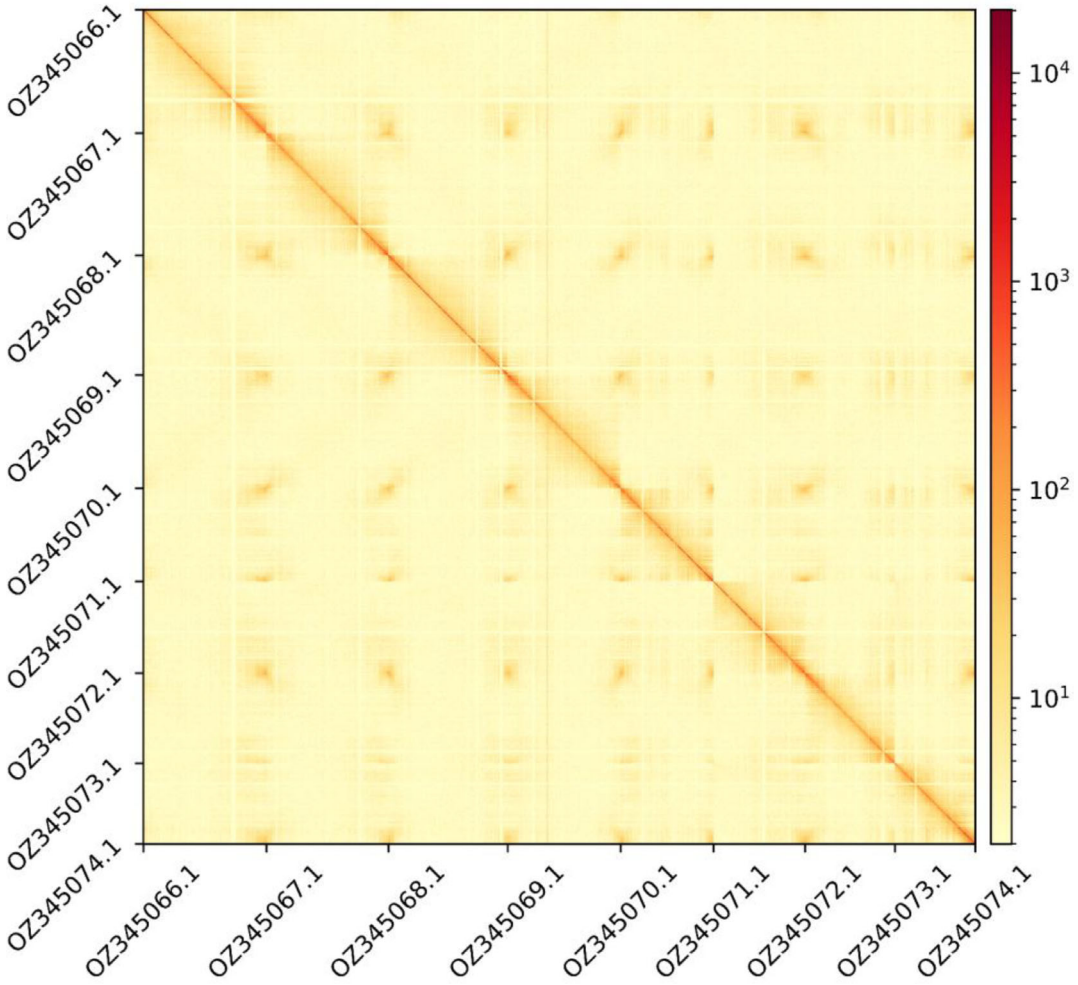


Figure 4. Hi-C contact map showing spatial interactions between regions in ddAlyUtri1.hap1 genome assembly. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 150 kb bin size for plotting.

Results

Genome assembly

The genome was assembled into two haplotypes using Hi-C phasing. Both haplotypes were curated to chromosome level (Table 1). Haplotype 1 had a total length of 1,012,219,214 bp in 228 scaffolds with a GC content of 39% (Figure 2), a contig N50 of 108,093,000 bp (L50 = 5) and a scaffold N50 of 135,936,682 bp (L50 = 4). This haplotype had 82 gaps, totalling 16.4 kb in size. The chloroplast genome was also assembled and had a total length of 153,111 bp.

Conserved single-copy gene content was assessed using the Eudicots odb10 database with BUSCO v5.8.0 (Manni et al., 2021). The hap1 assembly had 97.9% completeness (95.5% single-copy and 2.4% duplicated). 98.04% of reads' *k*-mers were present in the assembly (Figure 3), and the assembly had an average base accuracy Quality Value (QV) of 66.3 as calculated by Merqury (Rhie et al., 2020).

Chromosome-level scaffolds were named according to size, as confirmed by Hi-C data (Figure 4).

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.

Author contributions

CdG coordinated the project; RMW supervised the project; CdG and AF collected the species; AF identified the species; CdG sampled and preserved biological material and provided metadata; AsB, RM, TM, RAO, and THS provided support in sampling, shipping of biological material, metadata collection, and management; CN extracted DNA and CV extracted RNA, they prepared libraries, and performed sequencing under the supervision of CC; MAD edited and prepared reads for genome assembly; WSITOL Management, Samples and Laboratory Team and WSI Scientific Operations: Sequencing Operations extracted DNA, prepared libraries, and performed sequencing under the supervision of CH and MB; MS and TB performed genome assembly and curation; CB 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.

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 acknowledge the assembly reviewer Dom Absolon from the Wellcome Sanger Institute.

Literature cited

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit–interactive quality assessment of genome assemblies.** *G3: Genes, Genomes, Genetics.* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Chen S, Zhou Y, Chen Y, *et al.*: **fastp: an ultra-fast all-in-one FASTQ preprocessor.** *Bioinformatics.* 2018; **34**(17): i884–i890.
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

De Panis D: **Galaxy Workflow ERGA-BGE Genome Report ASM analyses (one-asm Long Reads + HiC).** 2024.
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

Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022. [Accessed 19 October 2022].

InfoFlora: 2026. Retrieved 24 August 2026.
Reference Source

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

Marhold K: **IAPT/IOPB chromosome data 1.** *Taxon.* 2006; **55**(2): 443–445.
Publisher Full Text

Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet. journal.* 2011; **17**(1): 10–12.
Publisher Full Text

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

POWO: *Plants of the World Online.* Kew: Facilitated by the Royal Botanic Gardens; 2026. Retrieved 24 August 2026.
Reference Source

Ranallo-Benavidez TR, Jaron KS, Schatz MC: **GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes.** *Nat. Commun.* 2020; **11**(1): 1432.
PubMed Abstract | Publisher Full Text | Free Full Text

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

Tvedte ES, Astashyn A, Sweeney D, *et al.*: **Rapid and sensitive detection of genome contamination at scale with FCS-GX.** *Genome Biol.* 2024; **25**.

Uliano-Silva M, Ferreira JGR, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | Publisher Full Text | Free Full Text

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

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