Check for updates

DATA NOTE

ERGA-BGE genome of *Androsace saussurei* Dentant, Lavergne, F. C. Boucher & S. Ibanez, a recently described plant from the rooftops of Europe

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

Sébastien Lavergne[1], Zoé Rosa[1], Chuan Peng[1-4], Rita Monteiro[5], Astrid Böhne[5], Thomas Marcussen[6,7], Torsten H. Struck[6], Rebekah A Oomen[6-10], Genomescope Sequencing Team, Alice Moussy[11], Corinne Cruaud[11], Karine Labadie[11], Emilie Téodori[12], Lola Demirdjian[12], Patrick Wincker[12], Pedro H Oliveira[12], Jean-Marc Aury[12], Vianey Paola Barrera Enriquez[13], Leanne Haggerty[13], Fergal Martin[13], Tom Brown[14,15]

[1]Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, LECA (Laboratoire d'Écologie Alpine), Grenoble, France
[2]CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan, 666303, China
[3]University of Chinese Academy of Sciences, Beijing, 100049, China
[4]Center of Plant Ecology, Core Botanical Gardens, Chinese Academy of Sciences, Mengla, Yunnan, 666303, China
[5]Leibniz Institute for the Analysis of Biodiversity Change - Museum Koenig Bonn, Bonn, 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, Canada
[9]Tjärnö Marine Laboratory, University of Gothenburg, Gothenburg, Sweden
[10]Centre for Coastal Research, University of Agder, Kristiansand, Norway
[11]Genoscope, Institut François Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, 2 Rue Gaston Crémieux, 91057, Evry, France
[12]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France
[13]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK
[14]Leibniz Institute for Zoo and Wildlife Research, Berlin, 10315, Germany
[15]Berlin Center for Genomics in Biodiversity Research, Berlin, 14195, Germany

v1 **First published:** 27 Nov 2025, **5**:368
https://doi.org/10.12688/openreseurope.21922.1
Latest published: 27 Nov 2025, **5**:368
https://doi.org/10.12688/openreseurope.21922.1

Abstract

The reference genome of *Androsace saussurei* (Ericales: Primulaceae: Primuloideae) provides a valuable resource for understanding the mechanisms of plant adaptations to the highest altitudes where flowering plants can thrive and the speciation mechanisms of plants dwelling on high alpine sky islands. The entirety of the genome

Open Peer Review

Approval Status ✓ ?

	1	2
version 1 27 Nov 2025	✓ view	? view

1. **Francesco Garassino**, Universität Zürich,

sequence was assembled into 20 contiguous chromosomal pseudomolecules. This chromosome-level assembly encompasses 413 Mb, composed of 253 contigs and 101 scaffolds, with contig and scaffold N50 values of 4.95 Mb and 21.2 Mb, respectively.

Keywords
Androsace saussurei, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, high alpine, cushion plant, Primulaceae, Ericales, Androsace de Saussure, Saussure's rock jasmine

Zürich, Switzerland

2. **Cheng-Yuan Zhou**, Fujian Agriculture and Forestry University (Ringgold ID: 12449), Fuzhou, China

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

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: Tom Brown (brown@izw-berlin.de)

Author roles: Lavergne S: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Rosa Z**: Investigation, Resources, Writing – Review & Editing; **Peng C**: Investigation, Resources, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Böhne A**: Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Marcussen T**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Struck TH**: Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Oomen RA**: Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Moussy A**: Investigation, Writing – Review & Editing; **Cruaud C**: Investigation, Writing – Review & Editing; **Labadie K**: Investigation, Writing – Review & Editing; **Téodori E**: Data Curation, Formal Analysis, Writing – Review & Editing; **Demirdjian L**: Data Curation, Formal Analysis, Writing – Review & Editing; **Wincker P**: Investigation, Supervision, Writing – Review & Editing; **Oliveira PH**: Investigation, Supervision, Writing – Review & Editing; **Aury JM**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Barrera Enriquez VP**: Data Curation, Formal Analysis, Writing – Review & Editing; **Haggerty L**: Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Martin F**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Brown T**: Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: Sampling was supported by a grant from LABEX OSUG@2020 (Investissements d'avenir -- ANR10 LBX56). 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. 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: © 2025 Lavergne S *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: Lavergne S, Rosa Z, Peng C *et al.* **ERGA-BGE genome of *Androsace saussurei* Dentant, Lavergne, F. C. Boucher & S. Ibanez, a recently described plant from the rooftops of Europe [version 1; peer review: 1 approved, 1 approved with reservations]** Open Research Europe 2025, **5**:368 https://doi.org/10.12688/openreseurope.21922.1

First published: 27 Nov 2025, **5**:368 https://doi.org/10.12688/openreseurope.21922.1

Introduction

Androsace saussurei is part of the plant family Primulaceae (Ericales) and is endemic to the highest summits of the Alps (Mont Blanc, Valais, Gran Paradiso, Vanoise) where it grows at the upper altitudinal limits of vascular plant life (from 2600 to 4300m a.s.l. in the Alps). The proposed species is one of three plant species recently described from the rooftops of Europe (Boucher *et al.*, 2021). This species was named after the pioneer, alpine scientist, Horace Bénédict de Saussure (1740–1799), who first reported this plant in 1788 from the Mont Blanc range, but which remained unnamed until recent scientific explorations.

These plants form very rare and particular 'cushion' life forms that are extremely long-lived (several hundreds of years) and play a key role in plant and insect diversity at these elevations, acting as a shelter for other species. This role as a keystone engineer species is essential to the maintenance of multi-trophic diversity in high alpine conditions that are otherwise adverse to many species. The genus Androsace, encompassing ca 173 species (POWO, 2024), is typical of high alpine and arctic environments of the Northern Hemisphere, and hence the genome of the study species constitutes a valuable resource for studying the mechanisms and scenarios of plant speciation and adaptation to the most extreme environments where flowering plants can thrive.

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 aim 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

Sébastien Lavergne, with help from Chuan Peng and Zoé Rosa, sampled one specimen of hermaphrodite *A saussurei.* from the Arête des Grandes Autannes, Le Tour, Chamonix, France on 7 July 2023. The species was identified by Sébastien Lavergne based on the recent description of the species (Boucher *et al.*, 2021). Sampling was performed under permit DDT-2017-1423 issued by the Préfecture de Haute Savoie (France). Sampling was performed via collection on the field, taken down through a one-hour hike, after which leaves and buds were picked from the plant and flash frozen using liquid nitrogen. Until DNA extraction, samples were preserved at -80C.

Vouchering information

Physical reference material for the sequenced specimen has been deposited in Grenoble Museum, France, herbarium code Gr (https://www.grenoble.fr/74-museum-de-grenoble.htm) under the accession number GR005037.

Frozen reference tissue material of leaf from the same individual is available at the Biobank INRAE-CNRGV, Plant Genomic Center, Toulouse, France (https://cnrgv.toulouse.inrae.fr/fr) under the voucher ID 2024-PS3939.

Genetic information

The estimated genome size, based on ancestral taxa, is 1.67 Gb with a monoploid number of 10 chromosomes, a haploid number of 20 and a ploidy of four. All information for this species was retrieved from Genomes on a Tree (Challis *et al.*, 2023).

We estimated the genome size and ploidy using PacBio HiFi data, analyzed with GenomeScope v2.0 and Smudgeplot v0.2.5, respectively. We report an estimated monoploid genome size of 237 Mb and a ploidy level of 4, tetraploid (Figure 1)

DNA/RNA processing

DNA was extracted from 1 g of leaves using a conventional CTAB extraction followed by purification using Qiagen Genomic tips (QIAGEN, MD, USA). A detailed protocol is available on protocols.io (https://dx.doi.org/10.17504/protocols.io.bp2l694yzlqe/v1). DNA fragment size selection was performed using Short Read Eliminator (PacBio, CA, USA). 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 usage.

RNA was extracted from leaves (50 mg) using the RNeasy Powerplant kit Plus Universal kit (Qiagen) following manufacturer instructions. Residual genomic DNA was removed with 6U of TURBO DNase (2 U/µL) (Thermo Fisher Scientific). Quantification was performed using a Qubit RNA HS Assay kit and integrity was assessed in a Bioanalyzer system (Agilent). RNA was stored at -80 °C.

Library preparation and sequencing

Long-read DNA libraries were prepared with the SMRTbell prep kit 3.0 following manufacturers' instructions and sequenced on a Revio system (PacBio). Omni-C libraries were prepared using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, CA, USA) after plant nuclei isolation. Briefly, flash-frozen young leaves (1 g) were cryoground in liquid nitrogen and pure nuclei were first isolated following the "High Molecular Weight DNA Extraction from Recalcitrant Plant Species" protocol described by Workman *et al.* (https://doi.org/10.1038/protex.2018.059). Once the nuclei had been isolated, the pellets were treated as mammalian cells and Omni-C libraries were prepared according to the Mammalian Cell Protocol for Sample Preparation v1.4. The final libraries were sequenced on an Illumina NovaSeq 6000 instrument (Illumina, San Diego, CA, USA) with 2x150 bp read length. In total 76.5 Gb PacBio HiFi and 35.4 Gb HiC data were sequenced to generate the assembly.

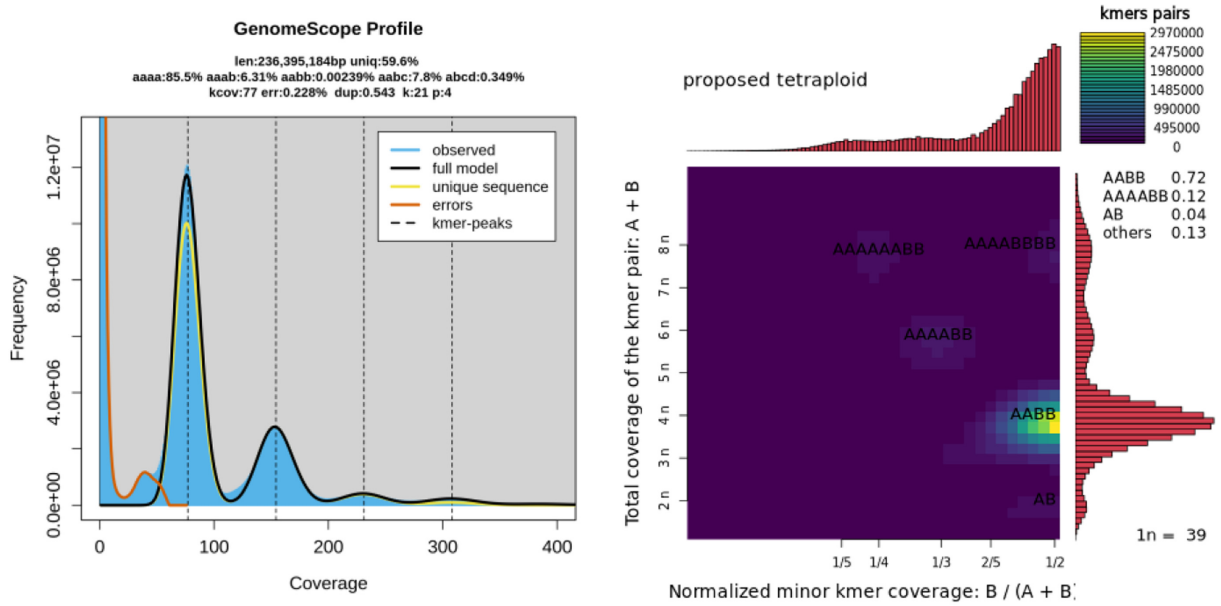


Figure 1. Kmer-based estimates of genome size, ploidy and heterozygosity. (Left) GenomeScope plot displaying the coverage of 31-mers (x-axis) against frequency x coverage of each 21-mer (y-axis). The black curve shows an inferred kmer distribution based on a ploidy of 4 and the inferred genome size (len), percentage of genome within homozygous (aaaa) or polymorphic regions (aaab, aabb, aabc, abcd), error rate (err) and unique sequence in the genome (uniq). The yellow curve shows the model based on unique 21-mers. The orange curve shows the curve based on the error 21-mers (Right) Smudgeplot displaying the normalised coverage (x-axis) against total coverage (y-axis) of each 21-mer in the sequencing dataset. Highlighted are elevated pairwise kmer coverage peaks, indicative of alternate ploidies. The probability of the best tested ploidy models are shown on the right, with the tetraploid model AABB showing the highest probability of 72%.

Genome assembly methods

The genome of *Androsace saussurei* was assembled using the Genoscope GALOP pipeline (https://workflowhub.eu/workflows/1200). Briefly, contigs were assembled using NextDenovo v2.5.1, followed by scaffolding using YaHS v1.2. The assembled scaffolds were then curated through manual inspection using PretextView v0.2.5 to remove any false joins and to incorporate sequences not automatically scaffolded into their respective locations within the chromosomal pseudomolecules. The mitochondrial and chloroplast genomes were assembled as single circular contigs using Oatk v1.0 and included in the released assembly (GCA_964267185.2).

As the genome is tetraploid, allelic duplications were not removed after the initial assembly. Instead, the two haplotypes were separated during manual curation, resulting in two chromosome-scale assemblies, each comprising 20 chromosomes (GCA_964267185.2 and GCA_964267195.1). Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (De Panis, 2024a).

Genome annotation methods

A gene set was generated using the Ensembl Gene Annotation system (Aken *et al.*, 2016), primarily by aligning publicly available short-read RNA-seq data from BioSamples: SAMN39954916, SAMN11835390, and SAMN11835391 to the genome. Gaps in the annotation were filled via protein-to-genome alignments of a select set of clade-specific proteins from UniProt (The UniProt Consortium, 2019), which had experimental evidence at the protein or transcript level. At each locus, data were aggregated and consolidated, prioritising models derived from RNA-seq data, resulting in a final set of gene models and associated non-redundant transcript sets. To distinguish true isoforms from fragments, the likelihood of each open reading frame (ORF) was evaluated against known metazoan proteins. Low-quality transcript models, such as those showing evidence of fragmented ORFs, were removed. In cases where RNA-seq data were fragmented or absent, homology data were prioritised, favouring longer transcripts with strong intron support from short-read data. The resulting gene models were classified into two categories: protein-coding, and long non-coding. Models that did not overlap protein-coding genes, and were constructed from transcriptomic data were considered potential lncRNAs. Potential lncRNAs were further filtered to remove single-exon loci due to their unreliability. Putative miRNAs were predicted by performing a BLAST search of miRBase (Kozomara *et al.*, 2019) against the genome, followed by RNAfold analysis (Gruber *et al.*, 2008). Other small non-coding loci were identified by scanning the genome with Rfam (Kalvari *et al.*, 2018) and passing the results through

Infernal (Nawrocki & Eddy, 2013). Summary analysis of the released annotation was performed using the ERGA-BGE Genome Report ANNOT Galaxy workflow (De Panis, 2024b).

Results

Genome assembly

The primary genome assembly has a total length of 413,403,326 bp in 101 scaffolds including the mitochondrial and chloroplast genomes (Figure 2 & Figure 3), with a GC content of 33.05%. The assembly has a contig N50 of 4,952,245 bp and L50 of 29 and a scaffold N50 of 21,172,409 bp and L50 of 10 with a total of 152 gaps. The single-copy gene content analysis using the Eudicots database with BUSCO (Manni *et al.*, 2021) resulted in 95.0% completeness (58.0% single and 37.0% duplicated). 63.3% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 53.6 as calculated by Merqury (Rhie *et al.*, 2020).

Genome annotation

The genome annotation consists of 28,787 protein-coding genes with an associated 43,854 transcripts, in addition to 6,389 non-coding RNA genes of various types (Table 1). Using the longest isoform per transcript, the single-copy gene content analysis using the eudicot_odb10 database with BUSCO (Manni *et al.*, 2021), run in protein mode, resulted in 90.4% completeness. Using the OMAmer Asterids database for OMArk (Nevers *et al.*, 2025) resulted in 90.49% completeness and 89.80% consistency (Table 2).

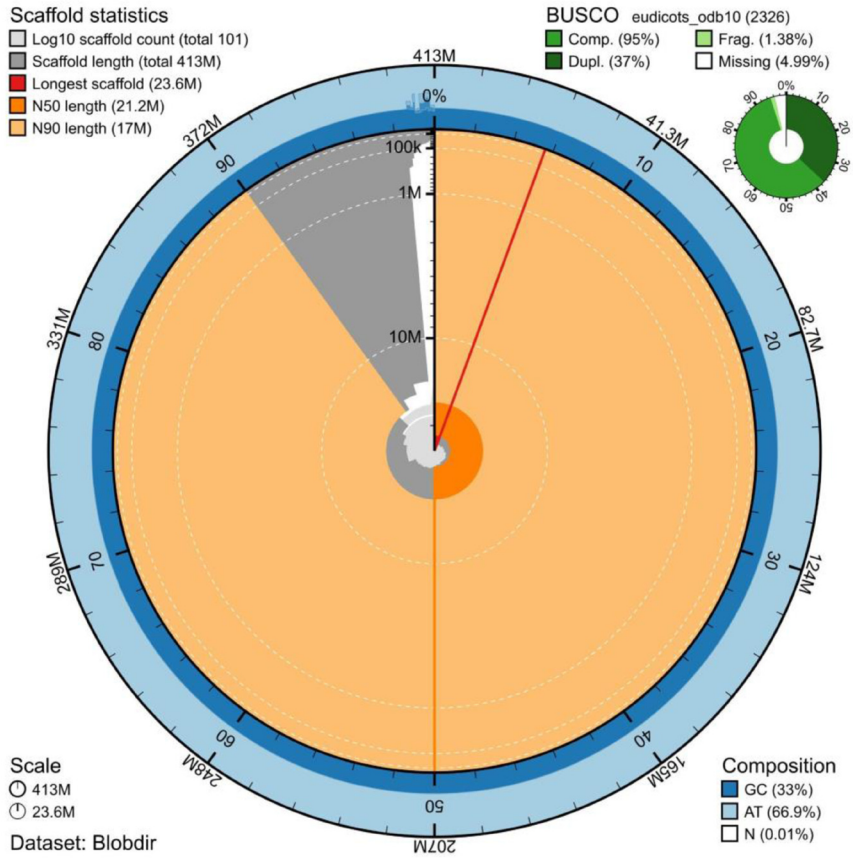


Figure 2. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 413,403,326 bp assembly including the mitochondrial genome. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (23,550,734 bp, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (21,172,409 and 17,031,950 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.

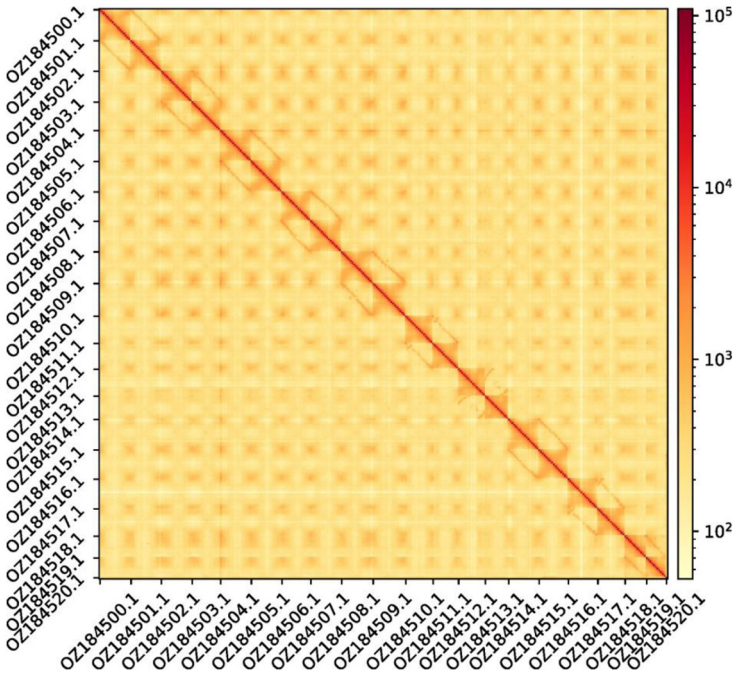


Figure 3. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 25 kb bin size for plotting.

Table 1. Statistics from assembled gene models.

	No. genes	No. transcripts	Mean* gene length (bp)	No. single-exon genes	Mean* exons per transcript
Protein-coding	28,787	43,854	3,528	743	5.3
lncRNA	1,832	1,872	1,044	7	2.6
snRNA	383	383	149	383	1.0
snoRNA	540	540	116	540	1.0
rRNA	1,183	1,183	149	1,183	1.0
tRNA	948	948	74	948	1.0
Other non-coding	1,503	1,503	86	1,503	1.0

*Combined categories show the range of the mean values

Table 2. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (eudicot_odb) and OMArk (Asterids).

	Complete	Singular	Duplicated	Fragmented	Missing
BUSCO	2,104 (90.4%)	1,610 (69.2%)	494 (21.2%)	74 (3.2%)	148 (6.4%)
OMArk	10,134 (90.49%)	5,829 (52.05%)	4,305 (38.44%)	-	1,064 (9.51%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	25,851 (89.80%)	1,945 (6.76%)	0 (0.00%)	991 (3.44%)	

Data availability

Androsace saussurei and the related genomic study were assigned to Tree of Life ID (ToLID) ddAndSaus6 and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB77232. The sample information is available at the following BioSample accessions: SAMEA114552830, and SAMEA115799864. The genome assembly is accessible from ENA under accession number GCA_964267185.1. Sequencing data produced as part of this project are available from ENA at the following accessions: ERX12733449, ERX12733450, and ERX12737191. 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/Androsace_saussurei/ddAndSaus6/. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/Androsace%20saussurei.

Author contributions

SL, CP and ZR collected the species, SL identified the species, SL sampled and preserved biological material and provided metadata, TM, RM, RO, THS and AsB provided support in sampling, shipping of biological material, metadata collection, and management, the GST extracted DNA, prepared libraries, and performed sequencing under the supervision of AM, CC, KL, PHO and PW; LD, JMA and ET performed genome assembly and curation,VPBE, LM and FM performed genome annotation, TB 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://doi.org/10.5281/zenodo.14611490

Acknowledgements

We acknowledge the co-descriptors of the species, Cédric Dentant, Florian Boucher and Sébastien Ibanez. We would like to acknowledge the assembly reviewer, Sarah Pelan, from the Wellcome Sanger Institute. The authors 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.

References

Aken BL, Ayling S, Barrell D, *et al.*: **The Ensembl gene annotation system.** *Database (Oxford).* 2016; **2016**: baw093.
PubMed Abstract | Publisher Full Text | Free Full Text

Boucher FC, Dentant C, Ibanez S, *et al.*: **Discovery of cryptic plant diversity on the rooftops of the Alps.** *Sci Rep.* 2021; **11**(1): 11128.
PubMed Abstract | Publisher Full Text | Free Full Text

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

De Panis D: **ERGA-BGE Genome Report ASM analyses (one-asm HiFi + HiC)**. WorkflowHub, 2024a.
Publisher Full Text

De Panis D: **ERGA-BGE Genome Report ANNOT analyses**. WorkflowHub, 2024b.
Publisher Full Text

Gruber AR, Lorenz R, Bernhart SH, *et al.*: **The Vienna RNA websuite.** *Nucleic Acids Res.* 2008; **36**(Web Server issue): W70–W74.
PubMed Abstract | Publisher Full Text | Free Full Text

Kalvari I, Nawrocki EP, Argasinska J, *et al.*: **Non-coding RNA analysis using the Rfam database.** *Curr Protoc Bioinformatics.* 2018; **62**(1): e51.
PubMed Abstract | Publisher Full Text | Free Full Text

Kozomara A, Birgaoanu M, Griffiths-Jones S: **miRBase: from microRNA sequences to function.** *Nucleic Acids Res.* 2019; **47**(D1): D155–D162.
PubMed Abstract | Publisher Full Text | Free 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.
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.
PubMed Abstract | Publisher Full Text

Nawrocki EP, Eddy SR: **Infernal 1.1: 100-fold faster RNA homology searches.** *Bioinformatics.* 2013; **29**(22): 2933–2935.
PubMed Abstract | Publisher Full Text | Free Full Text

Nevers Y, Warwick Vesztrocy A, Rossier V, *et al.*: **Quality assessment of gene repertoire annotations with OMArk.** *Nat Biotechnol.* 2025; **43**(1): 124–133.
PubMed Abstract | Publisher Full Text | Free Full Text

POWO: **Plants of the World Online**. Facilitated by the Royal Botanic Gardens, Kew, 2024; Retrieved 05 December 2024.
Reference Source

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

The UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res.* 2019; **47**(D1): D506–D515.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ?

Version 1

Reviewer Report 06 January 2026

https://doi.org/10.21956/openreseurope.23716.r65246

© **2026 Zhou C.** 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.

?

Cheng-Yuan Zhou
Fujian Agriculture and Forestry University (Ringgold ID: 12449), Fuzhou, Fujian, China

Sorry for the delay.
I have carefully examined the manuscript and the data. This study presents a high-quality reference genome of *Androsace saussurei* assembled using Illumina, PacBio HiFi, and Hi-C data. The dataset is substantial, and the manuscript is generally information-rich and well documented. However, a few minor issues should be addressed:
1. I agree with the authors' assembly and annotation workflow, and the amount of raw data appears sufficient. However, this is an allopolyploid tetraploid genome. Based on my experience, an assembly strategy using hifiasm followed by Hi-C scaffolding with HapHiC often yields improved results (and with relatively short runtime). If feasible, I would like to see a comparison between the current assembly and an assembly generated using a hifiasm + HapHiC pipeline.
2. I recommend that the authors annotate repetitive elements, as repeat composition is an important indicator of genome characteristics.
3. The Methods section could be described in greater detail. In particular, the authors should report the specific command lines and key parameters used, to facilitate reproducibility. This would substantially improve the readability and scientific rigor of the manuscript.

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?
Partly

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Plant genomics, systematics, taxonomy

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

Reviewer Report 02 January 2026

https://doi.org/10.21956/openreseurope.23716.r66209

© 2026 Garassino F. 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.

Francesco Garassino
Universität Zürich, Zürich, Switzerland

The Note by Lavergne et al. provides the scientific community with a valuable genomic resource from a lesser studied biotope.

The provided information on the recently discovered species justifies the sequencing and annotation of its genome. The methodology is clearly described and the essential assembly and annotation statistics are presented. I'd like to praise the Authors for including information on specimen availability.

While the Note is currently fully readable, I'd suggest a full proof-reading to improve the flow of some parts of the text (e.g., sentence "These plants form very rare and particular 'cushion' life forms that are extremely long-lived (several hundreds of years) and play a key role in plant and insect diversity at these elevations, acting as a shelter for other species.", while clear, comes as a bit of a surprise following the previous paragraph). Additional citations, if available, could be added to better circumstantiate the very interesting information provided in the Introduction.

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

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 ecophysiology, computational biology

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.