Check for updates

DATA NOTE

ERGA-BGE reference genome of *Carlina diae,* an endemic spineless thistle of Crete, Greece

[version 1; peer review: 3 approved with reservations]

Eleftheria Antaloudaki[1], Danae Karakasi[1,2], Manos Stratakis[1], Eleftherios Bitzilekis[1], Petros Lymberakis[1], Nikolaos Poulakakis[1,2], Tereza Manousaki[3], Astrid Böhne[4], Rita Monteiro[4], Rosa Fernández[5], Nuria Escudero[5], Genomescope Sequencing Team, Alice Moussy[6], Corinne Cruaud[6], Karine Labadie[6], Lola Demirdjian[7], Patrick Wincker[7], Pedro H. Oliveira[7], Jean-Marc Aury[7], Fergal Martin[8], Vianey Paola Barrera Enriquez[8], Leanne Haggerty[8], Chiara Bortoluzzi[9]

[1]Natural History Museum of Crete, School of Sciences and Engineering, Knossos Avenue, University of Crete, Heraklion, GR-71409, Greece
[2]Department of Biology, School of Sciences and Engineering, Vassilika Vouton, University of Crete, Heraklion, GR-70013, Greece
[3]Institute of Marine Biology Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, Crete, 70014, Greece
[4]Leibniz Institute for the Analysis of Biodiversity Change, Adenauerallee 127, Museum Koenig Bonn, Bonn, 53113, Germany
[5]Metazoa Phylogenomics Lab, Passeig marítim de la Barceloneta 37-49, Institute for Evolutionary Biology (CSIC-UPF)., Barcelona, 08003, Spain
[6]Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France
[7]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France
[8]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK
[9]SIB Swiss Institute of Bioinformatics, Amphipôle, Quartier UNIL-Sorge, Lausanne, 1015, Switzerland

v1 **First published:** 03 Feb 2026, **6**:34
https://doi.org/10.12688/openreseurope.22092.1
Latest published: 03 Feb 2026, **6**:34
https://doi.org/10.12688/openreseurope.22092.1

Abstract

Carlina diae (*Asteraceae*) is a rare Cretan endemic plant confined to limestone cliff crevices in four locations of eastern Crete, as well as Dia and the Dionysades islets. It is the only species of the *Carlina* genus lacking spiny structures, a trait that, together with its placement in the ancestral subgenus *Lyrolepis*, highlights its significance as a Tertiary relict and a representative of an ancient evolutionary lineage. Fewer than 1,000 mature individuals persist in fragmented subpopulations, each typically numbering fewer than 250 plants, all within Natura 2000 sites, which are protected areas covering Europe's most valuable and threatened species and habitats. The species is currently classified as Endangered on the IUCN Red List. Despite its restricted range, *C. diae* contributes substantially to ecosystem functioning. It

Open Peer Review

Approval Status ? ? ?

	1	2	3
version 1 03 Feb 2026	? view	? view	? view

1. **Nunzio D'Agostino**, University of Naples Federico II, Portici, Italy
2. **Francesco Garassino**, Universität Zürich, Zürich, Switzerland
3. **Cassandra Elphinstone**, University of British Columbia, Vancouver, Canada

Any reports and responses or comments on the

can stabilize shallow soils on rocky slopes and support the integrity of fragile montane habitats. Flowering during the dry summer months, its capitula provides critical nectar and pollen resources for pollinators, including solitary bees, thereby sustaining biodiversity during periods of resource scarcity. Unlike several species of the *Asteraceae* family with ornamental value, *C. diae* is not cultivated and is poorly represented in ex-situ collections, underscoring the urgency of conservation measures. Genomic studies, including full genome annotation, are expected to play a pivotal role in safeguarding its genetic diversity, informing effective management strategies, and improving understanding of its unique evolutionary history. The entirety of the genome sequence was assembled into 10 contiguous chromosomal pseudomolecules, 2 mitochondrial genomes, and one plastid genome. This chromosome-level assembly encompasses 4.2 Gb, composed of 456 contigs and 102 scaffolds, with contig and scaffold N50 values of 23.5 Mb and 416.6 Mb, respectively.

Keywords
Carlina diae, genome assembly, European Reference Genome Atlas, Biodiversity Genomics Europe, Earth Biogenome Project, Asteraceae, Cretan flora, endemic plant, Caduae, Καρλίνα του Δία

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: Chiara Bortoluzzi (chiara.bortoluzzi@sib.swiss)

Author roles: Antaloudaki E: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Karakasi D**: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Stratakis M**: Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Bitzilekis E**: Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Lymberakis P**: Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Poulakakis N**: Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Manousaki T**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Böhne A**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Monteiro R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Fernández R**: Methodology, Project Administration, Supervision, Writing – Review & Editing; **Escudero N**: Methodology, Project Administration, 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; **Wincker P**: Investigation, Supervision, Writing – Review & Editing; **Oliveira PH**: Investigation, Supervision, Writing – Review & Editing; **Aury JM**: Data Curation, Formal Analysis, Supervision, Writing – Review & Editing; **Martin F**: Data Curation, Formal Analysis, Writing – Review & Editing; **Barrera Enriquez VP**: Data Curation, Formal Analysis, Writing – Review & Editing; **Haggerty L**: Data Curation, Formal Analysis, 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. This work was supported by the Genoscope, the Commissariat à l'Energie 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 Antaloudaki E *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: Antaloudaki E, Karakasi D, Stratakis M *et al.* **ERGA-BGE reference genome of *Carlina diae,* an endemic spineless thistle of Crete, Greece [version 1; peer review: 3 approved with reservations]** Open Research Europe 2026, **6**:34 https://doi.org/10.12688/openreseurope.22092.1

First published: 03 Feb 2026, **6**:34 https://doi.org/10.12688/openreseurope.22092.1

Introduction

The genus *Carlina* L. (*Asteraceae*) comprises a group of thistle-like taxa, including annual, biennial, and perennial species, typically characterized by spiny leaves and distinctive capitula, with both ligulate and tubular florets surrounded by robust, often leaf-like, spiny bracts (Dimopoulos *et al*., 2013; Tutin, 1964).

Native to Europe, the Mediterranean Basin, and parts of Asia, several members of the genus have long attracted attention for their ornamental qualities and seeds of some *Carlina* species, remain commercially available for ornamental horticulture and landscaping. Among the members of this genus, *Carlina dia,* also known by its local name Καρλίνα του Δία, stands out due to its rarity and unique appearance - as it is the only species in the genus *Carlina* with no spines. This species is an endemic plant of Crete, Greece with a restricted and fragmented distribution, found in four locations in eastern Crete Island, in Dia islet and in Dionysades islet group. It is confined to crevices of hard limestone cliffs, usually near the coast and flowers during summer months (Grigoriadou *et al*., 2020). From an evolutionary perspective, it is considered a Tertiary relict and belongs to the small and ancestral subgenus *Lyrolepis* Meusel & Kästner (Grigoriadou *et al*., 2020). This phylogenetic position renders it an unusual representative of the genus and underscores its importance as a relic of an older evolutionary lineage.

Fewer than 1,000 mature individuals persist in highly fragmented subpopulations, each typically numbering fewer than 250 plants. All known subpopulations occur within the NATURA 2000 sites GR4310003 and GR4320006, and the species is formally protected under the Bern Convention (Appendix I) as well as the Greek Presidential Decree 67/1981. The International Union for Conservation of Nature (IUCN) currently classifies *C. diae* as Endangered, with an ongoing decline in population size, primarily driven by overgrazing and habitat degradation (Fournaraki *et al*., 2024).

Although *Carlina dia* is extremely localized, it plays a significant role in the functioning of the ecosystems it inhabits. Unlike most species of the genus, it lacks spiny structures, yet its perennial rosettes help stabilize shallow soils on rocky slopes and contribute to the maintenance of fragile montane habitats (Sardans & Peñuelas, 2013). The species' capitula, which flower during the dry summer months, provide critical nectar and pollen resources for a range of pollinators, including solitary bees and other insects that sustain local biodiversity (Herrera, 2024; Lucas *et al*., 2018). In addition, its growth form likely contributes to creating small microhabitats that support invertebrate assemblages, thereby enhancing habitat heterogeneity (Grigoriadou *et al*., 2020; Sardans & Peñuelas, 2013; Valli *et al*., 2021). Beyond these ecological functions, *C. diae* represents a unique evolutionary lineage whose loss would not only diminish the functional diversity of Cretan Mountain ecosystems but also erase a living relic of Tertiary floras.

Unlike other species with a history of ornamental use, *Carlina diae* is not cultivated and only a few individuals are maintained in ex-situ collections. Consequently, its conservation requires urgent and targeted measures. In this context, genomic studies—and particularly the comprehensive annotation of its genome—represent a crucial step toward safeguarding its genetic diversity, informing conservation strategies, and enhancing our understanding of its unique evolutionary history.

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

On 20th November 2023, Eleftheria Antaloudaki sampled leaves of one specimen of *Carlina diae* (hermaphrodite monoecious), which were identified using the Flora Europaea and the Atlas of the Aegean Flora (Strid, 2016). The specimen was collected originally on 9th May 2023 by Manolis Avramakis at the Gulf of Panagia at Dia islet and was moved and transplanted in the small botanic garden of the Natural History Museum of Crete at Knossou premises, Heraklion, Crete, Greece. Sampling was performed under Presidential Decree 67/1981 issued by the Greek Government. Approximately 10 g of fresh leaves were cut off the individual and were immediately flash frozen in liquid nitrogen. The material was stored in -80 °C until DNA extraction.

Vouchering information

Physical reference material for the here sequenced specimen has been deposited in the Herbarium of the Botany Division of the Natural History Museum of Crete https://www.nhmc.uoc.gr/en/departments/botany under accession number NHMC 42.13475.

Frozen leaf material is available from the same individual at the tissue collection of the Genomics and Genetic Resources Division of the Natural History Museum of Crete https://www.nhmc.uoc.gr/en/departments/genomics under voucher ID NHMC 42.13475.

Genetic information

The estimated genome size, estimated by Genomes on a Tree (GoaT) by ancestral state reconstruction, is 3.75 Gb. This is a diploid genome with a haploid number of 10 chromosomes (2n=20). The entire *Carlina* genus is uniformly diploid, with no known polyploid species within the genus. All information for this species was retrieved from Genomes on a Tree (Challis *et al*., 2023).

DNA/RNA processing

DNA was extracted from 1 g of leaves using a conventional CTAB extraction followed by a commercial purification using Qiagen Genomic tips (Qiagen). A detailed protocol is available on protocols.io (Vacherie *et al.*, 2022). 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 usage.

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). Two Omni-C libraries were prepared using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, CA, USA) after plant nuclei isolation. Briefly, flash-frozen leaves (1.5 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.* (2018). 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 X Plus instrument (Illumina, San Diego, CA, USA) with 2 × 150 read length. In total 61× PacBio HiFi and 210× 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 Hifiasm v0.21.0-r686 (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. This assembly was scaffolded using YaHS v1.2.2 (Zhou *et al.*, 2023) and assembled scaffolds were then curated through manual 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. 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).

Genome annotation methods

A gene set was generated using the Ensembl Gene Annotation system on a previously released version of the reference genome (GCA_965177975.1). (Aken *et al.*, 2016), primarily by aligning publicly available short-read RNA-seq data from BioSamples SAMEA112287425 and SAMN16450721 to the genome. Gaps in the annotation were filled via protein-to-genome alignments of a select set of clade-specific proteins from (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 plant 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).

Results

Genome assembly

The genome assembly has a total length of 4,151,721,093 bp in 102 scaffolds, including two mitogenomes and one plastid genome (Figure 1 & Figure 2), with a GC content of 39.3%. The assembly has a contig N50 of 23,495,317 bp and L50 of 55 and a scaffold N50 of 416,577,109 bp and L50 of 5. The assembly has a total of 354 gaps, totalling 36.4 kb in cumulative size. The single-copy gene content analysis using the Embryophyta database with BUSCO v5.8.0 (Manni *et al.*, 2021) resulted in 97.9% completeness (91.1% single and 6.8% duplicated). 98% of reads k-mers were present in the assembly and the assembly has a base accuracy Quality Value (QV) of 75.5 as calculated by Merqury (Rhie *et al.*, 2020).

Genome annotation

The genome annotation consists of 27,296 protein-coding genes with associated 37,392 transcripts, in addition to 27,739 non-coding genes (Table 1). Using the longest isoform per transcript, the single-copy gene content analysis using the embryophyta_odb10 database with BUSCO resulted in 88.8% completeness. Using the OMAmer Viridiplantae-v2.0.0.h5 database for OMArk (Nevers *et al.*, 2025) resulted in 94.5% completeness and 92.0% consistency (Table 2).

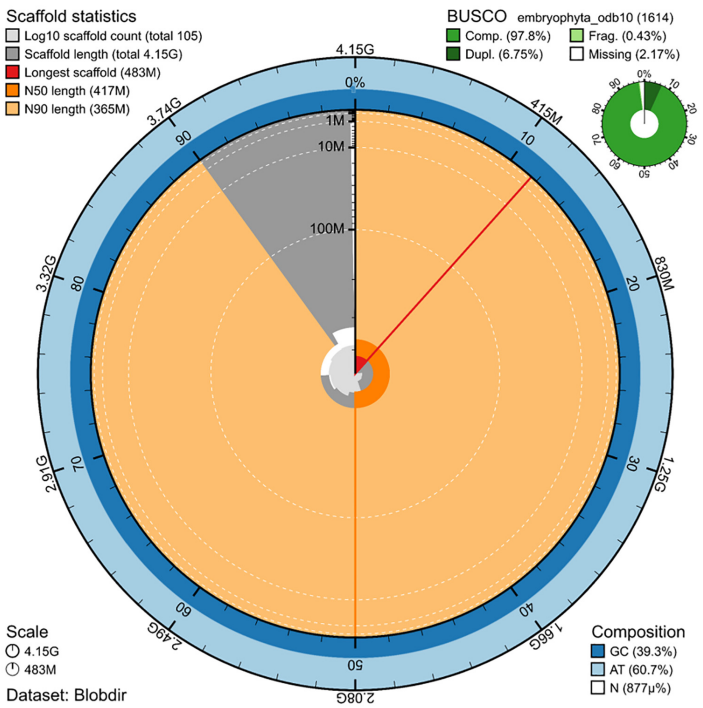


Figure 1. Snail plot summary of assembly statistics. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 4,151,721,093 bp assembly including the two mitochondrial genomes. The distribution of sequence lengths is shown in dark grey, with the plot radius scaled to the longest sequence present in the assembly (483 Mb, shown in red). Orange and pale-orange arcs show the scaffold N50 and N90 sequence lengths (416,577,109 bp and 365,390,074 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.

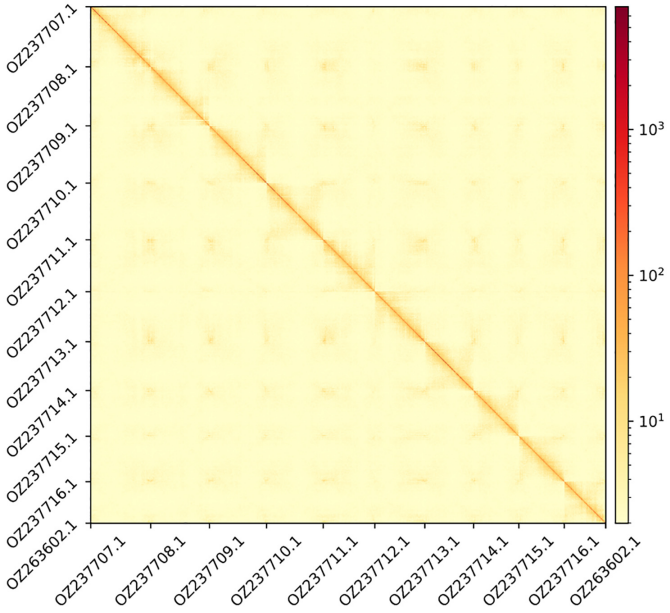


Figure 2. Hi-C contact map showing spatial interactions between regions of the genome. The diagonal corresponds to intra-chromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 400 kb bin size for plotting. The Hi-C contact map shows the 10 chromosomes and the plastid genome (GenBank accession: OZ263602.1).

Table 1. Statistics from assembled gene models.

	No. genes	No. transcripts	Mean gene length (bp)	No. single-exon genes	Mean exons per transcript
mRNA	27,296	37,392	4,619	1,689	5.4
pseudogene	0.00	0.00	0.00	0.00	0.00
snoRNA	13,205	13,205	106	13,205	1.0
lncRNA	5,658	5,963	1,254	640	2.2
snRNA	305	305	145	305	1.0
rRNA	7,778	7,778	203	7,778	1.0
tRNA	793	793	75	793	1.0
Other ncRNA	6,879	6,879	68 – 100	6,879	1.0 – 1.0

Table 2. Annotation completeness and consistency scores calculated by BUSCO run in protein mode (embryophyta_odb10) and OMArk (Viridiplantae-v2.0.0.h5).

	Complete	Single copy	Duplicated	Fragmented	Missing
BUSCO	1,433 (88.8%)	1,359 (84.2%)	74 (4.6%)	104 (6.4%)	77 (4.8%)
OMArk	10,586 (94.5%)	7,773 (69.4%)	2,813 (25.1%)	-	612 (5.5%)
	Consistent	**Inconsistent**	**Contaminants**	**Unknown**	
OMArk	25,101 (92.0%)	722 (2.6%)	0.0 (0.0%)	1,473 (5.4%)	

Data availability

Carlina diae and the related genomic study were assigned to Tree of Life ID (ToLID) 'daCarDiae1' and all sample, sequence, and assembly information are available under the umbrella BioProject PRJEB84159. The sample information is available at the following BioSample accessions: SAMEA115717571 and SAMEA115717572. The genome assembly is accessible from ENA under accession number GCA_965177975.2 and the annotation is 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: ERX14170105, ERX14170108, ERX14170109, ERX14170716, ERX14170717, ERX14170718, ERX14170719, ERX14170720, ERX14170721, ERX14170722, and ERX14170723. 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/Carlina_diae/daCarDiae1. Further details and data about the project are hosted on the ERGA portal at https://portal.erga-biodiversity.eu/data_portal/538483.

Author contributions

DK and RM coordinated the project; EA collected the species; EA identified the species; EA and DK sampled and preserved biological material and provided metadata; AsB, RM, RF, NE, MS, EB, PL, NP, and TM 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 and JMA performed genome assembly and curation under the supervision of JMA; FM, VPBE, and LH performed genome annotation; 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. 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

The authors would like to thank Manolis Avramakis, technician of the Botany Division of the Natural History Museum of Crete, who collected, transplanted the specimen from the wild, and maintains the botanic garden of NHMC, where Carlina diae is transplanted. 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, Fernando Cruz from the Centro Nacional de Análisis Genómico (CNAG).

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

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

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

Dimopoulos P, Raus T, Bergmeier E, *et al.*: **Vascular plants of Greece: an annotated checklist.** 2013.
Reference Source

Fournaraki C, Bazos I, Choreftakis M: **Carlina diae.** The IUCN Red List of Threatened Species 2024. 2024.

Grigoriadou K, Sarropoulou V, Krigas N, *et al.*: **GIS-facilitated effective propagation protocols of the endangered local endemic of crete *Carlina diae* (Rech. f.) Meusel and A. Kástner (Asteraceae): serving ex situ conservation needs and its future sustainable utilization as an ornamental.** *Plants (Basel).* 2020; **9**(11): 1465.
PubMed Abstract | Publisher Full Text | Free 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

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.
Reference Source

Herrera MC: **Refrigerated flowers in the torrid Mediterranean summer.** *Ecology.* 2024; **105**(3): e4268.
PubMed Abstract | Publisher 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

Lucas A, Bodger O, Brosi JB, *et al.*: **Floral resource partitioning by individuals within generalised hoverfly pollination networks revealed by DNA metabarcoding.** *Sci Rep.* 2018; **8**(1): 5133.
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

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

Sardans J, Peñuelas J: **Plant-soil interactions in Mediterranean forest and shrublands: impacts of climatic change.** *Plant Soil.* 2013; **365**(1–2): 1–33.
PubMed Abstract | Publisher Full Text | Free Full Text

Strid A: **Atlas of the Aegean Flora. Part 2: maps.** *Englera.* 2016; **33**: 1–878.
Reference Source

Tutin TG: **Flora Europaea: Alismataceae to Orchidaceae (Monocotyledones).** 1964.
Reference Source

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

Vacherie B, Labadie K, Falentin C: **HMW DNA extraction for long read sequencing using CTAB.** 2022.
Publisher Full Text

Valli AT, Koumandou VL, Iatrou G, *et al.*: **Conservation biology of threatened Mediterranean chasmophytes: the case of *Asperula naufraga* endemic to Zakynthos Island (Ionian Islands, Greece).** *PLoS One.* 2021; **16**(2): e0246706.
PubMed Abstract | Publisher Full Text | Free Full Text

Workman R, Timp W, Fedak R, *et al.*: **High molecular weight DNA extraction from recalcitrant plant species for third generation sequencing.** 2018.
Publisher Full Text

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

Open Peer Review

Current Peer Review Status: ? ? ?

Version 1

Reviewer Report 08 April 2026

https://doi.org/10.21956/openreseurope.23908.r69920

© **2026 Elphinstone 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.

?

Cassandra Elphinstone
University of British Columbia, Vancouver, British Columbia, Canada

Summary
The manuscript “ERGA-BGE reference genome of Carlina diae, an endemic spineless thistle of Crete, Greece” presents a chromosomes level assembly for Carlina diae, a rare Tertiary relict endemic to Crete, Greece. Using PacBio HiFi and Hi-C sequencing, the authors assemble a 4.2 Gb genome into 10 chromosomes, along with plastid and mitochondrial genomes.
The authors present a clear rationale for sequencing the genome of this endangered, Tertiary relict. The manuscript is generally well written, and the figures and tables are clear. However, there are some details that should be included to make the study fully reproducible.
Suggested revisions:
Intro
It would be helpful for the readers to redefine what the NATURA 2000 sites are in the introduction. Currently, this is only explained in the abstract.
Ensure consistent spelling of the species name (diae vs dia)
In this sentence “yet” is not the correct term – maybe split into two sentences. “Unlike most species of the genus, it lacks spiny structures, yet its perennial rosettes help stabilize shallow soils on rocky slopes and contribute to the maintenance of fragile montane habitats”
In the next paragraph, you should clarify that you are referring to other species in this genus with a history of ornamental use – otherwise this sentence implies Carlina diae is an ornamental but not cultivated.

It would be helpful for readers to include a photo of the species somewhere in the manuscript.
Methods
The ERGA’s sequencing strategy at the beginning of the methods section is unnecessary. It talks about ONT and Illumina PE sequencing which was not done for this assembly. Remove this to not mislead readers.
Rearrange the sampling description to clarify the reference individual was first moved from the wild to a botanical garden and then sampled in the botanical garden. Normally manuscripts do not mention specific people’s names in the sampling information.
Refer to “Physical reference material” as “a herbarium voucher from the sequenced individual” to

differentiate it from the frozen material described later.
Under “Genetic information”, move the results about the expected genome size and ploidy from GoaT to the results. The methods should only focus on where this information was obtained and how.
Under the DNA extraction section, clarify the weight is for fresh frozen tissue. It would also be helpful to state how long the HMW DNA was stored at 4C.
Under the “Genome assembly methods” section, remove the term “briefly” for Hifiasm and maybe clarify that the Genoscope GALOP pipeline runs the programs described (Hifiasm, YaHS, etc) – if I’m understanding this correctly.
The gene annotation methods are the least reproducible based on the current manuscript description. What tools were used to align RNAseq data? How do they compare with using a program like BRAKER? The total gene counts and OMArk scores are good but the BUSCO Embryophyta score is about 10% below the assembly BUSCO score. I would also double check the BUSCO details in both sections to make them consistent in phrasing.
Could you support this statement by any other studies that have also assumed this? “Models that did not overlap protein-coding genes and were constructed from transcriptomic data were considered potential lncRNAs.”
Results
I would not include the stats for the contig level assembly – just include the final stats for the manually, curated, published assembly. Also, how were the organelle genomes assembled? Does GALOP do this automatically – some more information should be included in the methods.
What is meant by two mitogenomes? The methods for this are not clear in the methods section.
Final summary
It would be helpful for readers at the end of the manuscript to have a short section outlining future uses and applications for this assembly. For instance, can this genome and annotation help us understand the genetic basis of the spines in other species in the Carlina? How can this genome help with the conservation of this endangered species?
Figures and tables
The figures and tables are clear with well described captions. Figure 1 presents the genome statistics but not the distribution of the annotation etc. across the 10 pseudochromosomes.
An additional figure that would be helpful to include would show the gene density (and maybe transposable element density) across the ten chromosomes. It would also be helpful for the reader to include a photo of the species somewhere in 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?

No

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

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, bioinformatics, Arctic alpine plants

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 12 March 2026

https://doi.org/10.21956/openreseurope.23908.r69917

© 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

This Data Note by Antaloudaki *et al.* is well-written and outlines well the sequencing and annotation of the genome of an interesting species. The genome sequence and this Data Note will certainly be a useful resource for the plant sciences community.

However, the Methods section needs to be revised in order to ensure reproducibility. I agree with the comments of Reviewer 1 and join them in asking the authors to provide additional information on the tools and parameters employed during the assembly and annotation steps. Alternatively, or additionally, if possible, the authors could provide a documented version of the pipeline or set of scripts they used to perform assembly and annotation.

I would also like to ask to clarify or edit the "2 mitochondrial genomes" statement, as it is essentially incorrect in my opinion.

I would also encourage the authors to include a (set of) picture(s) of the species and its habitat into the Note.

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?
No

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Plant ecophysiology, bioinformatics.

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 13 February 2026

https://doi.org/10.21956/openreseurope.23908.r69457

© 2026 D'Agostino N. 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.

?

Nunzio D'Agostino
University of Naples Federico II, Portici, Italy

The manuscript by Antaloudak et al. presents the genome assembly of *Carlina diae* (Asteraceae), a rare and endangered Cretan endemic species of high conservation and evolutionary interest. Despite its restricted distribution, *C. diae* plays a meaningful ecological role within its native habitats.

The authors report a chromosome-level genome assembly spanning 4.2 Gb, organized into 10 chromosomal pseudomolecules, together with the mitochondrial and plastid genomes. This genomic resource has the potential to support conservation strategies and contribute to a deeper understanding of the species' evolutionary history.

The manuscript is generally well written. The Introduction clearly outlines the rationale for generating the presented resource, and the study objectives are well defined and logically structured.

However, I have several concerns that should be addressed before the manuscript can be considered for indexing (see my detailed comments below). In particular, important methodological details are missing, which limits the reproducibility and transparency of the study. Moreover, results concerning the organelle genomes are either insufficiently described or entirely absent, despite being mentioned elsewhere in the manuscript. Addressing these issues would substantially strengthen the rigor, clarity, and overall completeness of the work.

MAJOR

It is not appropriate to refer to "two mitochondrial genomes." The authors likely mean that they assembled two alternative mitochondrial genome conformations, which arise from the high recombinational activity of the mitochondrial genome across direct and inverted repeat elements. Such recombination can generate structurally distinct isoforms (subgenomic molecules or alternative configurations) rather than representing two separate mitochondrial genomes.

The manuscript should clarify this point and adopt more accurate terminology, explicitly stating whether the assemblies correspond to alternative structural forms produced by recombination across repeat sequences. This distinction is important to avoid conceptual misunderstandings regarding mitochondrial genome organization.

In the *Gene Assembly Methods* section, several important details regarding the parameters used for each tool are missing. These should be explicitly reported to ensure full reproducibility of the analyses. Key information such as software versions, command-line options, threshold values, and any deviations from default settings must be clearly specified. Without these details, it is difficult for readers to accurately replicate or critically assess the assembly workflow.

Additionally, the manuscript does not mention the tools or strategies employed for assembling cytoplasmic organelle genomes. Given their distinct genomic features and the frequent use of specialized assembly approaches for organellar DNA, this information should be included. Clarifying whether dedicated tools were used, how organellar reads were identified and handled, and how assembly quality was evaluated would significantly strengthen the methodological transparency of the study.

A similar concern applies to the *Genome Annotation Methods* section, where critical methodological details are lacking. For example, the authors state that RNA-seq reads were aligned to the genome, but they do not specify which alignment tool was used, nor do they report the software version, parameter settings, or whether default options were applied. These details are essential to ensure reproducibility and to allow readers to properly evaluate the robustness of the annotation pipeline.

Likewise, the BLAST analyses are insufficiently described. The e-value threshold and any additional filtering criteria (e.g., percentage identity, coverage cutoffs, database version) are not reported. Such parameters can substantially influence annotation outcomes and should be clearly indicated.

Furthermore, there is no mention of the tools or workflows used for the annotation of cytoplasmic organelle genomes (e.g., mitochondrial and plastid genomes). Given the availability of dedicated annotation pipelines for organellar genomes and their distinct structural and functional features, the absence of this information represents a significant methodological gap. The authors should clarify how organellar genomes were annotated, including the software, databases, and parameter settings employed.

Overall, a more comprehensive and transparent description of the annotation procedures is necessary to ensure methodological rigor and reproducibility.

The Results section lacks a description of the cytoplasmic organelle genomes. If the authors choose to mention these genomes in the manuscript, it would be appropriate to provide a clear account of the results obtained. This should include relevant assembly features (e.g., genome size, structure, gene content), as well as any notable characteristics identified. Including this information would ensure completeness and improve the overall coherence of the study.

Finally, to the best of my knowledge, a Data Note should include a dedicated section outlining how the generated datasets can be reused by the scientific community. This aspect is currently missing from the manuscript. Including a clear statement on potential applications and possible reuse

scenarios would significantly enhance the value and impact of the work, while aligning it with the standard expectations for this type of publication.

MINOR:

In the Abstract, I suggest replacing the phrase *"full genome annotation"* with *"including genome sequencing and annotation."* The latter wording is more precise and avoids potential ambiguity or overstatement.

At the end of the Introduction, I suggest removing the expression *"comprehensive annotation,"* as certain important components—such as repetitive elements—do not appear to have been included in the annotation workflow.

In Table 1, I suggest removing the row reporting pseudogenes, as it does not appear to provide informative or meaningful insight in its current form.

I would suggest including a photograph of the species under investigation, as this would help readers better appreciate its morphology and ecological context. While not essential, such an addition would enhance the overall presentation 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?
No

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics and bioinformatics

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.