A Chromosome-Scale Genome Assembly and Annotation of the Antarctic Pearlwort *Colobanthus quitensis* (Caryophyllaceae)

Eduardo Castro-Nallar, Claudio Meneses, Amy E Zanne, Katarzyna Fudala, Robert Józef Bialik, Sergio Guajardo-Leiva, Claudia Egas, Marco A Molina-Montenegro, Cristóbal Galbán-Malagón, Ricardo Rozzi, Patricio Tapia-Reyes

A Chromosome-Scale Genome Assembly and Annotation of the Antarctic Pearlwort *Colobanthus quitensis* (Caryophyllaceae)

Eduardo Castro-Nallar[1,2,3*#], Claudio Meneses[4,5*#], Amy E. Zanne[6], Katarzyna Fudala[7], Robert Józef Bialik[7], Sergio Guajardo-Leiva[1,2], Claudia Egas[1,2], Marco A. Molina-Montenegro[1,2,8], Cristóbal Galbán-Malagón[9,10,11,12], Ricardo Rozzi[3], Patricio Tapia-Reyes[13,14*]

[1]Centro de Ecología Integrativa, Universidad de Talca, Campus Talca, Talca, Chile.

[2]Instituto de Ciencias Biológicas, Universidad de Talca, Campus Talca, Talca, Chile.

[3]Cape Horn International Center, Universidad de Magallanes, Teniente Muñoz 166, Puerto Williams 6350000, Chile.

[4]Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago 8331150, Chile.

[5]Departamento de Fruticultura y Enología, Facultad de Agronomía e Ingeniería Forestal, Pontificia Universidad Católica de Chile, Santiago 7820436, Chile.

[6]Cary Institute of Ecosystem Studies, Millbrook, New York, USA.

[7]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland.

[8] Centro de Investigación en Recursos Naturales y Sustentabilidad (CIRENYS), Universidad Bernardo O'Higgins, Avenida Viel #1497, Santiago, Chile.

[9]Centro de Genómica Ecología y Medio Ambiente (GEMA), Universidad Mayor, Santiago de Chile, Chile.

[10]Institute for Environment, Florida International University, Miami, FL, USA.

[11]Data Observatory Foundation, Santiago de Chile, Chile.

[12]Centro Tecnológico de Innovación en Envases LABEN-CHILE, Santiago de Chile, Chile.

[13]Escuela de Biotecnología, Facultad de Ciencias, Universidad Santo Tomás, Chile.

[14]Facultad de Medicina Veterinaria y Agronomía, campus Providencia, Universidad de Las Américas, Chile.

[#]Authors contributed equally.

[*]Corresponding authors.

Eduardo Castro-Nallar, ecastron@utalca.cl

Claudio Meneses, claudio.meneses@uc.cl

Patricio Tapia-Reyes, patricio.tapia.bioinfo@gmail.com

Abstract (250 words)

Colobanthus quitensis (Caryophyllaceae) is one of only two vascular plants native to Antarctica and a model species for studying adaptation to extreme environments. We report a chromosome-scale reference genome generated from PacBio HiFi and Hi-C sequencing. The final assembly spans 887.1 Mb and resolves into 40 chromosome-scale scaffolds (2n = 80), with a scaffold N50 of 20.4 Mb and a maximum scaffold length of 35.1 Mb. Assembly completeness was high, with 99.0% of embryophyta BUSCOs recovered, the majority as duplicated (95.7%), consistent with the species' putative tetraploid origin. The genome exhibits a GC content of 36.3%. We further assembled the complete chloroplast genome (151,314 bp; 113 genes) and a draft mitochondrial assembly comprising two partial contigs totaling ~289 kb. Repeat annotation identified 69.9% of the genome as repetitive, with long terminal repeat retrotransposons as the dominant class (34.8%), primarily of the Ty1/Copia and Gypsy/DIRS1 superfamilies. A total of 40,762 protein-coding genes were predicted, of which 95% were functionally annotated. Combined with transcriptomic resources, this genome enables functional annotation and comparative studies. This is the first chromosome-scale assembly for an Antarctic angiosperm and a resource for investigating polyploidy, genome organization, and adaptive evolution in polar environments. The *C. quitensis* genome will also support ecological and applied studies, including the development of crops resilient to cold and other abiotic stresses.

Keywords: Chromosome-scale genome assembly, Caryophyllaceae, Polyploidy, Antarctic angiosperm, *Colobanthus quitensis*

Significance (150 words)

Colobanthus quitensis is one of only two flowering plants native to Antarctica and a model for understanding how vascular plants establish and persist in extreme polar environments. We present the first chromosome-scale reference genome for this species, resolving its genome into 40 chromosome-level scaffolds (2n = 80). This assembly enables studies of polyploidy, genome organization, and chromatin architecture in a high-latitude angiosperm. With 99.0% BUSCO completeness and strong contiguity metrics, the genome supports comparative research within Caryophyllaceae and across polar taxa. It also enables investigation of the molecular basis of cold tolerance and stress adaptation. The *C. quitensis* genome also supports applied research on crop resilience and on the mechanisms of plant survival under extreme conditions.

Introduction

The flora of Antarctica has low species richness with only two native angiosperms, *Deschampsia antarctica* Desv. (Poaceae) and *Colobanthus quitensis* (Kunth.) Bartl. (Caryophyllaceae). Caryophyllaceae is a large angiosperm family, with ~3,000 species, primarily Holarctic in distribution but found worldwide (Harbaugh et al. 2010). *Colobanthus quitensis*, also known as Antarctic Pearlwort, exhibits a pan-American and Antarctic distribution, ranging from México to the Antarctic Peninsula (Convey et al. 2011; Moore 1970). Molecular work has revealed *C. quitensis* as a recent migrant to Antarctica, probably in the late Pleistocene (< 1 Mya; Biersma et al. 2020), and most studies have centered on the ecophysiology and adaptive traits of this plant, e.g., cold, drought, and UV tolerance.

C. quitensis exhibits sexual and asexual reproduction, though the former is the primary mode of establishment in natural populations (Giełwanowska et al. 2011; Zúñiga et al. 2008). Antarctic populations of *C. quitensis* form dense cushions and thick amphistomatic leaves that aid with water conservation and cold tolerance (Ramírez et al. 2024). Additionally, they exhibit seed heteromorphism, characterized by dark brown seeds with higher germination and salinity tolerance, thereby aiding establishment under stress (Ontivero et al. 2024). Physiologically, Antarctic populations of *C. quitensis* are better adapted to freezing and cold temperatures, are more efficient at photosynthesis, and possess antioxidant, photoprotective, and osmoprotective mechanisms (Cho et al. 2018; Gianoli et al. 2004; Min et al. 2025). It is believed that many of these traits are regulated by fungal endophytes, epigenetic modifications, and stress-induced gene expression (Acuña-Rodríguez et al. 2024; Barrera et al. 2020; Cho et al. 2018; Egas et al. 2025; Hereme et al. 2020). The molecular underpinnings of these traits are largely unknown, relying primarily on marker genes and reference-free *de novo* gene expression analyses.

Colobanthus quitensis is tetraploid (2C ~ 2.0 pg) with a known diploid population in Conguillío National Park, Chile (Cuba-Díaz et al. 2016; Pascual-Díaz et al. 2020) (2C ~ 1.0 pg). Its chloroplast genome has been sequenced and contains 112 genes (78 protein-coding, 30 tRNA, and four rRNA), with a genome structure highly congruent with that of *C. apetalus* (Kang et al. 2015). *C. quitensis de novo* transcriptome has been reported with more than 47,000 functionally annotated transcripts (Cho et al. 2018). *C. quitensis* populations have high genetic differentiation and low gene flow, where isolation by distance shapes genetic structure (Biersma et al. 2020; Cuba-Díaz et al. 2017; Koc et al. 2018).

We present the reference genome assembly of *Colobanthus quitensis* from the Maritime Antarctic region. This chromosome-level assembly is highly contiguous, with most chromosomes resolved from telomere to telomere, and includes annotated gene models. It will facilitate the study of trait variation within *C. quitensis* populations, help identify genes responsible for adaptation to extreme environments, and improve the precision of experiments assessing changes in gene expression under different conditions.

Results and Discussion

Genome Assembly

The assembly resolved into 40 chromosome-level scaffolds (93.8% of total length; Table 1; Figure 1A), spanning 887.1 Mb (consistent with the flow-cytometry estimate of 1C ~1 pg) with a scaffold N50 of 20.4 Mb and a maximum scaffold of 35.1 Mb. Assessment against the Embryophyta OrthoDB v12 dataset (n = 2,026 genes) yielded 99.0% BUSCO completeness (C:99.0% [S:3.3%, D:95.7%], F:0.3%, M:0.6%); most recovered genes (95.7%) were classified as duplicated, a pattern diagnostic of the species' tetraploid origin (2n = 80). K-mer completeness reached 98.4%, and the gap density of 2,075 gaps/Gbp represents a 10-fold improvement over the prior unpublished scaffold-level assembly (KOPRI_Coqu_10Xv1; Table 1), which achieved a scaffold N50 of only 2.3 Mb and 22,587 gaps/Gbp. This assembly, POLARIX, meets Earth BioGenome Project quality tier 6.7.Q30 (Table 1). Telomere identification using Quartet revealed that 29 chromosomes possess both telomeres, 10 chromosomes contain one telomere, and one chromosome lacks detectable telomeric sequences (Figure S5).

Repeat Annotation

Repeat annotation identified 69.9% of the *C. quitensis* genome as repetitive, with LTR retrotransposons constituting the dominant class at 34.8% of the total assembly (Table 1; Table S3). Ty1/Copia (21.6%) and Ty3/Gypsy (12.3%) dominated this class; LINEs and DNA transposons contributed 5.6% and 3.6%, respectively, with 26% unclassified by RepeatModeler (Table S3). The predominance of LTR retrotransposons in *C. quitensis* is consistent with patterns observed in other angiosperm genomes, in which the Gypsy and Copia superfamilies, organized into phylogenetically conserved lineages across eudicots (Neumann et al. 2019), are the principal drivers of genome-size variation and chromosomal restructuring. Whole-genome duplication can trigger retrotransposon activation by relaxing epigenetic silencing, with LTR eruptions documented in autopolyploids following chromosome doubling (He et al. 2022). Furthermore, the extreme abiotic conditions characteristic of Maritime Antarctica, sustained subzero temperatures, intense UV irradiance, and repeated freeze-thaw cycles, are recognized activators of stress-responsive retrotransposons in plants (Papolu et al. 2022). Ongoing polar environmental pressures may therefore continue to shape the transposable element content of *C. quitensis* independently of its polyploid origin.

Gene Functional Annotation

Gene prediction yielded 40,762 protein-coding gene models, of which 95.0% received functional annotations from eggNOG-mapper, including GO terms for 24,789 genes, PFAM domain assignments for 36,082 genes, and functional descriptions for 38,724 genes (Table 1; Table S4). BUSCO assessment of predicted proteins against the Embryophyta OrthoDB v12 dataset yielded 93.3% completeness, with 79.3% of recovered genes classified as duplicated, a pattern that mirrors the assembly-level BUSCO signal and is consistent with the retention of homeologous gene copies across the two subgenomes of this tetraploid species. OMArk assessment of Pentapetalae-conserved Hierarchical Orthologous Groups (HOGs) corroborated these results, with 87.2% of conserved gene families represented in the annotation (Table 1). The gene count of ~40,700 is broadly comparable to other polyploid Caryophyllaceae: the chromosome-level genome of *Gypsophila paniculata*, which has undergone two rounds of WGD, was annotated with approximately 86,000 gene models (Li et al. 2022) and the 54,342 models in the reference genome KOPRI_Coqu_10Xv1 (Table 1). Furthermore, while our annotation shows a higher duplication rate (79.3% in

BUSCO) compared to the reference (53.3%), this difference reflects our stringent filtering pipeline, which reduced the initial 51,156 candidates (Table S4) to 40,762 high-confidence models to minimize false positives. This conservative approach effectively removed fragmented sequences, as evidenced by increases in mean gene length (from 3,361 to 3,897 bp) and mean CDS length (from 1,203 to 1,372 bp). This reduction did not compromise the biological completeness of the annotation. While the raw annotation showed higher BUSCO (C:97.7%) and OMArk (C:93.49%) scores, our filtered set maintained robust levels (BUSCO C:93.3%; OMArk C:87.17%). The high proportion of duplicated genes (79.3% in BUSCO) across both sets supports the validity of the retained homeologous copies in this polyploid species and yields a high-confidence dataset for downstream evolutionary analyses. This pattern is relevant to cold adaptation: ancient WGD events in Caryophyllaceae have been linked to gene family expansions enriched for ubiquitination and stress-response functions, with duplicated copies in cushion plant lineages, including *Colobanthus*, proposed as contributors to the colonization of cold climatic niches (Feng et al. 2024).

Genome Architecture

Interchromosomal synteny analysis revealed that most chromosomes maintain connections with at least two others, reflecting retention of ancestral genomic architecture consistent with one or more WGD events in the lineage (Table S5; Figure 1B). Seven chromosomes displayed connections with more than three partners (Table S5), suggesting complex reorganization following polyploidization. Syntenic blocks were predominantly collinear, indicating that large-scale chromosomal rearrangements have been limited since the most recent WGD, with gene order broadly preserved across subgenomes. These findings are consistent with the documented paleopolyploid history of Caryophyllaceae, in which at least 15 putative ancient WGD events have been inferred from transcriptomic data, including events predating the family-level divergence (Feng et al. 2024), and multiple WGDs have been associated with shifts into colder climatic occupancies across the Caryophyllales (Smith et al. 2017; Yang et al. 2015). Reads not assigned to *C. quitensis* yielded 113 high-quality metagenome-assembled genomes (MAGs) spanning 13 bacterial phyla, serving as a companion resource for studying the plant-associated microbiome (Table S8; Figures S3–S4).

Chloroplast and mitochondrial genomes

Assembly of organellar genomes from HiFi reads yielded one complete plastid genome and two partial plastid contigs, as well as two partial mitochondrial contigs. The complete plastid genome spans 151,314 bp and encodes 113 genes (Figure S1, Table S6), closely matching the size and gene content of other Caryophyllaceae plastomes, which range from approximately 150,224 bp in *Silene capitata* to over 152,000 bp in *Lychnis wilfordii* (Kang et al. 2017). Regarding the additional partial plastid and mitochondrial contigs, their significantly lower read coverage (~50X) compared to the complete plastid genome (~1000X) indicates that these sequences might be assembly artifacts or exogenous contaminants rather than genuine structural variants of the *C. quitensis* genome. The mitochondrial assembly comprised two partial contigs totaling ~289 kb and encoding six protein-coding genes each (Figure S2, Table S7). The fragmented mitochondrial assembly is consistent with the highly dynamic and recombination-prone mitochondrial genome architecture characteristic of angiosperms, and particularly of Caryophyllaceae, which exhibit elevated rates of mitochondrial gene loss and genome restructuring compared to most other plant families (Sloan et al. 2010). Consequently, these partial mitochondrial contigs should be interpreted with caution, as their reduced coverage suggests they may represent non-target sequences captured during assembly.

1

Genome assembly	Current study	Previous reference
Assembly Name	POLARIX	KOPRI_Coqu_10Xv1 (PRJNA1100813)
Assembly Level	Chromosome	Scaffolds
Assembly size (bp)	887,064,685	905,644,381
Number of contigs	2,678	48,598
Number of scaffolds	626	28,143
Longest scaffold	35,143,692	13,269,416
GC content (%)	36.3	32.64
EBP Quality Standard	6.7.Q30	5.6.Q59
Contigs NG50	801,095	49,700
Scaffolds NG50	21,037,107	2,310,145
Gaps/Gbp (1811 gaps)	2,075.16	22,587.23
Consensus quality QV	29.75	58.65
K-mer completeness	98.4%	95.5%
Assembly: BUSCO (embryophyta_odb12: 2,026 BUSCO genes)	C:99.0%[S:3.3%,D:95.7%],F:0.3%,M:0.6%	C:98.1%[S:8.1%,D:89.9%],F:1.0%,M:0.9%
Genome masked	69.98%	
LINEs(L1/CIN4)	5.58%	-
LTR (Gypsy/DIRS1 + Ty1/Copia)	34.84%	-
DNA transposon	3.55%	-
Unclassified Repeats	26%	-
Number of proteins-coding genes	40,762	54,342
genes with associated GO terms	24,789	-
Genes with PFAM annotation	36,082	-
Genes with description	38,724	-
Annotation: BUSCO	C:93.3%[S:14.1%,D:79.3%],F:0.8%,M	90.0%[S:36.8%,D:53.3%],F:5.9%,M:4

(embryophyta_odb12: 2,026 BUSCO genes)	:5.9%	.0%
Annotation: OMArk (Pentapetalae: OMA / All.Jul2023, 9983 conserved HOGs)	C:87.17%[S:12.64%, D:74.53%],M:12.83%	C:93.27%[S:21.01%, D:72.26%],M:6.73%

Table 1. Assembly statistics and quality assessment of the *Colobanthus quitensis* genome.

Assembly metrics for the primary contigs generated with hifiasm (PacBio HiFi reads) and the chromosome-scale scaffolds obtained after Hi-C–based scaffolding with YaHS. Assembly length, contiguity metrics (N50, L50, N80, L80), and GC content are shown. BUSCO assessment was performed against the Embryophyta 12 dataset (n = 2,026 genes), indicating 99.0% completeness in the final assembly, with only 0.6% of genes missing. The scaffolded assembly spans ~887 Mb across 626 scaffolds (≥ 10 kb), with an N50 of 20.36 Mb, demonstrating near-complete, chromosome-scale resolution.

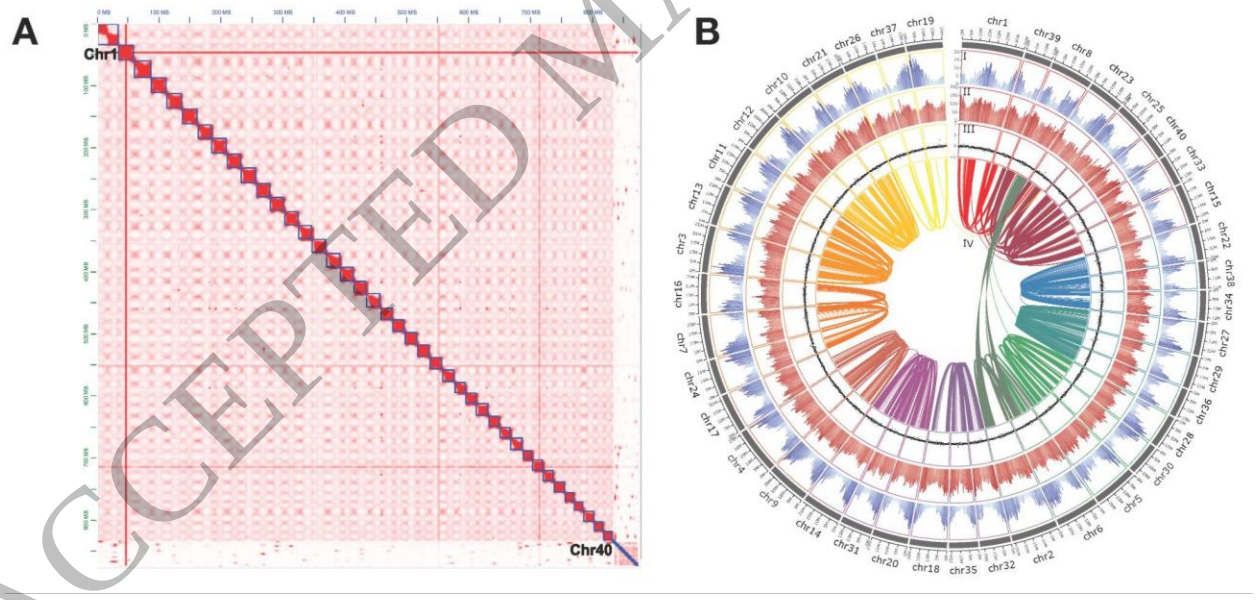

Figure 1. Chromosome-scale Hi-C contact map of the *Colobanthus quitensis* genome assembly.

A: Hi-C interaction heatmap of the *C. quitensis* genome assembled from PacBio HiFi long reads and scaffolded with Hi-C data. Interaction frequencies are shown on a red–white scale, with stronger contacts along the diagonal. The map reveals 40 discrete chromosome-scale scaffolds, sorted and named by length from chr1 to chr40, each visible

as a distinct square along the diagonal, indicating successful reconstruction of the haploid chromosome set. Off-diagonal contact patterns reflect long-range chromatin interactions within and between chromosomes. The uniformity and sharpness of the contact map demonstrate the high contiguity and accuracy of the final assembly, spanning ~900 Mb.

B: Chromosome-scale ideogram of the *Colobanthus quitensis* genome assembly. Chromosomes were ordered based on the number of intrachromosomal connections from MCScanX. I: gene density shown in blue, II: transposon density shown in red, III: GC skew. IV: intrachromosomal synteny relationships calculated from MCScanX; ribbons of the same color represent clusters of chromosomes that share the highest number of connections among themselves compared to the rest. Ideogram generated in 100 Kbp windows.

Materials and Methods

Sample Collection and DNA Extraction

Plant material was collected in King George Island, South Shetland Archipelago, Antarctica (Puchalski Point; Arctowski Polish Antarctic Station; -62.162, -58.471). DNA extraction was carried out from a single individual of *Colobanthus quitensis*, collected in a UV-radiated sterile plastic bag and frozen until processing. Sample collection and export were authorized by the Instituto Antártico Chileno (INACH), in accordance with the Protocol on Environmental Protection to the Antarctic Treaty. DNA extraction was conducted using the QIAGEN Genomic DNA Purification procedure.

PacBio Library and Sequencing

DNA samples were quantified on a Qubit 4.0 (ThermoFisher) and assessed on a Femto Pulse (Agilent) prior to library preparation. QC-passing DNA was sheared on a MegaRuptor 3 (Diagenode) to 15–18 kb, then prepared with the SMRTbell Prep Kit 3.0 (Pacific Biosciences). Adaptor-ligated libraries were size-selected with AMPure PB beads or on a BluePippin (Sage Science) and sequenced on a Revio SMRT cell (25M) for 24 h.

Dovetail Omni-C library preparation and sequencing

Dovetail Omni-C libraries were prepared following the manufacturer's protocol: in-nucleus formaldehyde crosslinking, DNase I digestion, ligation of a biotinylated bridge adapter, proximity ligation, crosslink reversal, and DNA purification. Sequencing libraries were prepared with NEBNext Ultra enzymes and Illumina-compatible adapters, enriched after

streptavidin pull-down of biotin-containing fragments, and sequenced on an Illumina HiSeqX at 30X coverage.

Read Filtering and Decontamination

Omni-C Hi-C reads were processed using Trim Galore! v0.6.10 (Krueger 2019) with parameters "-q 28 --length 100 --paired" to remove adapters and low-quality sequences. To isolate *C. quitensis* reads from potential contaminants, a decontamination blacklist was generated. Taxonomic identification of contaminants was performed using the sendsketch.sh script from BBMap v39.45, which identified eight bacterial and arthropod contaminant genomes (Table S2). Reference genomes were retrieved from NCBI and concatenated with the existing *C. quitensis* reference genome (GCA_041430805.1) to create a combined reference file. PacBio HiFi reads were then aligned against this combined reference using minimap2 v2.3r1287 (Li 2018) with options "-p 0.7 -N 50 --cs". Reads whose primary alignment, highest identity, and CIGAR score corresponded to *C. quitensis* (GCA_041430805.1) scaffolds were selected from the resulting SAM file.

These selected reads were further validated using Kraken2 (Wood et al. 2019) v2.17.0 against the PlusPFP database (updated October 2025) with the option ·--minimum-hit-groups 5". Reads classified within the Viridiplantae taxonomic group (taxid: 33090) were extracted using extract_kraken_reads.py from KrakenTools (Lu et al. 2022) v1.2.1 with the parameters "-t 33090 --include-children". Viridiplantae and unclassified reads formed the final *C. quitensis* dataset; remaining reads were assigned to the plant-associated microbiome.

Genome Assembly and Scaffolding

Filtered Hi-C and HiFi reads were assembled using Mabs (Schelkunov 2023) v2.28 with the parameters "--local_busco_dataset eudicots_odb10.2024-01-08.tar.gz --number_of_busco_orthogroups all --ploidy 2 --genome_size 1G". Reads from the filtered Omni-C library were mapped to the assembled primary contigs using bwa (Li 2013) v0.7.19-r123 and subsequently filtered using pairtools v0.3.0, following the Omni-C protocol published online. The resulting data were utilized in YAHS (Zhou et al. 2022) v1.1.2 for scaffolding. Finally, the scaffolds were manually corrected using Juicer, JuicerTools, and Juicebox. (Durand et al. 2016). Additionally, telomeres were identified using Quartet v.1.2.5r3 with the TeloExplorer function (Lin et al. 2023).

Genome Annotation

Transposable and repetitive elements were annotated using reasonaTE (TransposonUltimate pipeline; Riehl et al. 2022). Repetitive elements were classified using RepeatClassifier from RepeatModeler (Flynn et al. 2020) v2.0.7 and used to mask the genome with RepeatMasker (Tempel 2012) v4.2.1 ("-xsmall -nolow -no_is -norna -qq -gff -e rmblast"). The masked genome was annotated using BRAKER v3.0.8 supported by 29 mRNA-seq datasets from various *C. quitensis* experiments (Table S9), 3,185 Viridiplantae proteins from Swiss-Prot, and annotated proteins from 23 Caryophyllales species from NCBI (Table S10). The BRAKER (Gabriel et al. 2024) annotation was filtered using AGAT v1.5.0 to remove mono-exonic genes lacking functional or PFAM annotations, genes with mRNA lengths < 150 bp, and secondary isoforms. Prior to BUSCO assessment of the predicted proteome, a single representative protein per locus was retained by keeping only the longest isoform using agat_sp_keep_longest_isoform.pl (AGAT v1.5.0). Functional annotation used eggNOG (Cantalapiedra et al. 2021)-mapper v2.1.12 against EggNOG 5.0 (Table S11). Genome assembly completeness was assessed with Merqury v1.3 and BUSCO v6.0.0 (embryophyta OrthoDB v12), while the completeness of the filtered protein-coding annotation was evaluated using BUSCO v6.0.0 (embryophyta OrthoDB v12) and OMARk (Nevers et al. 2024) v0.3.0. Organellar genomes were assembled using PMAT (Bi et al. 2024) v1.5.3 and annotated with GeSeq (Tillich et al. 2017) v2.03, using *Silene latifolia* (NC_014487) and *C. quitensis* (NC_028080) as mitochondrial and plastid references, respectively.

Synteny Analysis

An all-against-all BLASTp (Camacho et al. 2009) v2.17.0 alignment ("-evalue 1e-10 -outfmt 6 -num_alignments 5") of annotated proteins was analyzed with MCScanX (Wang et al. 2012) v1.0.0 ("-b 1") to identify intra-chromosomal syntenic blocks.

Metagenomic Assembly and Binning

HiFi reads that were not classified as belonging to the *C. quitensis* genome were assembled using hifiasm-meta v0.3.1. (Feng et al. 2022). The resulting contigs were processed with the HiFi-MAG-Pipeline (Portik et al. 2024) (from pb-metagenomics-tools) to generate high-quality Metagenome-Assembled Genomes (MAGs). Only MAGs exhibiting >70% completeness, <10% contamination, and a maximum of 20 contigs were retained for further analysis.

Data availability

The raw sequencing data, assembled genome, and organelle sequences have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA1457453. The genome assembly is available at GenBank accession JBXXPJ000000000. Annotated gene models, repeat libraries, and organelle assemblies are accessible via Zenodo (https://doi.org/10.5281/zenodo.19741749). Supporting transcriptome datasets are available under BioProject PRJNA388703.

References

Acuña-Rodríguez IS, et al. 2024. Fungal endophyte symbionts enhance plant adaptation in Antarctic habitats. Physiologia Plantarum. 176(6). doi:10.1111/ppl.14589

Barrera A, et al. 2020. Fungal Endophytes Enhance the Photoprotective Mechanisms and Photochemical Efficiency in the Antarctic *Colobanthus quitensis* (Kunth) Bartl. exposed to UV-B Radiation. Frontiers in Ecology and Evolution. 8. doi:10.3389/fevo.2020.00122

Bi C, et al. 2024. PMAT: an efficient plant mitogenome assembly toolkit using low-coverage HiFi sequencing data. Horticulture Research. 11(3). doi:10.1093/hr/uhae023

Biersma EM, et al. 2020. Multiple late-Pleistocene colonisation events of the Antarctic pearlwort *Colobanthus quitensis* (Caryophyllaceae) reveal the recent arrival of native Antarctic vascular flora. Journal of Biogeography. 47(8):1663-1673. doi:10.1111/jbi.13843

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10(1). doi:10.1186/1471-2105-10-421

Cantalapiedra CP, et al. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution. 38(12):5825-5829. doi:10.1093/molbev/msab293

Cho SM, et al. 2018. Comparative transcriptome analysis of field- and chamber-grown samples of *Colobanthus quitensis* (Kunth) Bartl, an Antarctic flowering plant. Scientific Reports. 8(1). doi:10.1038/s41598-018-29335-4

Convey P, Hopkins DW, Roberts SJ, Tyler AN. 2011. Global southern limit of flowering plants and moss peat accumulation. Polar Research. 30(1):8929. doi:10.3402/polar.v30i0.8929

Cuba-Díaz M, Cerda G, Rivera C, Gómez A. 2016. Genome size comparison in *Colobanthus quitensis* populations show differences in species ploidy. Polar Biology. 40(7):1475-1480. doi:10.1007/s00300-016-2058-z

Cuba-Díaz M, et al. 2017. Phenotypic variability and genetic differentiation in continental and island populations of *Colobanthus quitensis* (Caryophyllaceae: Antarctic pearlwort). Polar Biology. 40(12):2397-2409. doi:10.1007/s00300-017-2152-x

Durand NC, et al. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Systems. 3(1):95-98. doi:10.1016/j.cels.2016.07.002

Egas C, et al. 2025. Fungal endophytes modulate the negative effects induced by Persistent Organic Pollutants in the antarctic plant *Colobanthus quitensis*. Physiologia Plantarum. 177(1). doi:10.1111/ppl.70079

Feng X, Cheng H, Portik D, Li H. 2022. Metagenome assembly of high-fidelity long reads with hifiasm-meta. Nature Methods. 19(6):671-674. doi:10.1038/s41592-022-01478-3

Feng K, et al. 2024. The link between ancient whole-genome duplications and cold adaptations in the Caryophyllaceae. American Journal of Botany. 111(8). doi:10.1002/ajb2.16350

Flynn JM, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences. 117(17):9451-9457. doi:10.1073/pnas.1921046117

Gabriel L, et al. 2024. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Research. 34(5):769-777. doi:10.1101/gr.278090.123

Gianoli E, et al. 2004. Ecotypic Differentiation in Morphology and Cold Resistance in Populations of *Colobanthus quitensis* (Caryophyllaceae) from the Andes of Central Chile and the Maritime Antarctic. Arctic, Antarctic, and Alpine Research. 36(4):484-489. doi:10.1657/1523-0430(2004)036[0484:EDIMAC]2.0.CO;2

Giełwanowska I, et al. 2011. Biology of generative reproduction of *Colobanthus quitensis* (Kunth) Bartl. from King George Island, South Shetland Islands. Polish Polar Research. 32(2):139-155. doi:10.2478/v10183-011-0008-6

Harbaugh DT, et al. 2010. A New Lineage-Based Tribal Classification of the Family Caryophyllaceae. International Journal of Plant Sciences. 171(2):185-198. doi:10.1086/648993

He J, et al. 2022. An Eruption of LTR Retrotransposons in the Autopolyploid Genomes of *Chrysanthemum nankingense* (Asteraceae). Plants. 11(3):315. doi:10.3390/plants11030315

Hereme R, et al. 2020. Fungal Endophytes Exert Positive Effects on *Colobanthus quitensis* Under Water Stress but Neutral Under a Projected Climate Change Scenario in Antarctica. Frontiers in Microbiology. 11. doi:10.3389/fmicb.2020.00264

Kang Y, et al. 2015. The complete chloroplast genome of Antarctic pearlwort *Colobanthus quitensis* (Kunth) Bartl. (Caryophyllaceae). Mitochondrial DNA Part A. 27(6):4677-4678. doi:10.3109/19401736.2015.1106498

Kang JS, Lee BY, Kwak M. 2017. The complete chloroplast genome sequences of *Lychnis wilfordii* and *Silene capitata* and comparative analyses with other Caryophyllaceae genomes. PLOS ONE. 12(2):e0172924. doi:10.1371/journal.pone.0172924

Koc J, et al. 2018. Range-wide pattern of genetic variation in *Colobanthus quitensis*. Polar Biology. 41(12):2467-2479. doi:10.1007/s00300-018-2383-5

Krueger F. 2019. Trim Galore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files. Available from: https://github.com/FelixKrueger/TrimGalore

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34(18):3094-3100. doi:10.1093/bioinformatics/bty191

Li F, et al. 2022. The chromosome-level genome of *Gypsophila paniculata* reveals the molecular mechanism of floral development and ethylene insensitivity. Horticulture Research. 9. doi:10.1093/hr/uhac176

Lin Y, et al. 2023. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Horticulture research*, *10*(8), p.uhad127.

Lu J, et al. 2022. Metagenome analysis using the Kraken software suite. Nature Protocols. 17(12):2815-2839. doi:10.1038/s41596-022-00738-y

Min K, et al. 2025. Transcriptomic responses of Antarctic plants to *in situ* warming: uncovering molecular mechanisms behind physiological adjustments. Annals of Botany. 136(3):529-542. doi:10.1093/aob/mcaf108

Moore D. 1970. Studies in *Colobanthus quitensis* (Kunth) Bartl. and *Deschampsia antarctica* Desv. II. Taxonomy, distribution and relationships. British Antarctic Survey Bulletin. 23:63--80.

Neumann P, Novák P, Hoštáková N, Macas J. 2019. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mobile DNA. 10(1). doi:10.1186/s13100-018-0144-1

Nevers Y, et al. 2024. Quality assessment of gene repertoire annotations with OMArk. Nature Biotechnology. 43(1):124-133. doi:10.1038/s41587-024-02147-w

Ontivero Y, et al. 2024. Preliminary assessment of seed heteromorphism as an adaptive strategy of *Colobanthus quitensis* under saline conditions. Scientific Reports. 14(1). doi:10.1038/s41598-024-82381-z

Papolu PK, et al. 2022. Retrotransposons: How the continuous evolutionary front shapes plant genomes for response to heat stress. Frontiers in Plant Science. 13. doi:10.3389/fpls.2022.1064847

Pascual-Díaz JP, et al. 2020. Genome size constancy in Antarctic populations of *Colobanthus quitensis* and *Deschampsia antarctica*. Polar Biology. 43(9):1407-1413. doi:10.1007/s00300-020-02699-y

Portik DM, et al. 2024. Highly accurate metagenome-assembled genomes from human gut microbiota using long-read assembly, binning, and consolidation methods. bioRxiv. doi:10.1101/2024.05.10.593587

Ramírez CF, et al. 2024. Ecophysiology of Antarctic Vascular Plants: An Update on the Extreme Environment Resistance Mechanisms and Their Importance in Facing Climate Change. Plants. 13(3):449. doi:10.3390/plants13030449

Riehl K, Riccio C, Miska EA, Hemberg M. 2022. TransposonUltimate: software for transposon classification, annotation and detection. Nucleic Acids Research. 50(11):e64-e64. doi:10.1093/nar/gkac136

Schelkunov MI. 2023. Mabs, a suite of tools for gene-informed genome assembly. BMC Bioinformatics. 24(1). doi:10.1186/s12859-023-05499-3

Sloan DB, et al. 2010. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. BMC Evolutionary Biology. 10(1):274. doi:10.1186/1471-2148-10-274

Smith SA, et al. 2017. Disparity, diversity, and duplications in the Caryophyllales. New Phytologist. 217(2):836-854. doi:10.1111/nph.14772

Tempel S. 2012. Using and Understanding RepeatMasker. Methods in Molecular Biology. 29-51. doi:10.1007/978-1-61779-603-6_2

Tillich M, et al. 2017. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Research. 45(W1):W6-W11. doi:10.1093/nar/gkx391

Wang Y, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research. 40(7):e49-e49. doi:10.1093/nar/gkr1293

Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biology. 20(1). doi:10.1186/s13059-019-1891-0

Yang Y, et al. 2015. Dissecting Molecular Evolution in the Highly Diverse Plant Clade Caryophyllales Using Transcriptome Sequencing. Molecular Biology and Evolution. 32(8):2001-2014. doi:10.1093/molbev/msv081

Zhou C, McCarthy SA, Durbin R. 2022. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 39(1). doi:10.1093/bioinformatics/btac808

Zúñiga GE, Zamora P, Ortega M, Obrecht A. 2008. Short Note: Micropropagation of Antarctic *Colobanthus quitensis*. Antarctic Science. 21(2):149-150. doi:10.1017/s0954102008001570