

Reference Genome Assembly of a Tetraploid Accession of the Tuber Crop *Tropaeolum tuberosum*

Daniela Ruiz-Mateus ^{1,†}, Ina Scheffler ^{2,†}, María del Pilar Márquez-Cardona ¹, Thomas Greb ^{2,3}, Wilson Terán ¹, Pascal Hunziker ^{2,3,*}

¹Biología de Plantas y Sistemas Productivos, Departamento de Biología, Pontificia Universidad Javeriana, Bogotá, Colombia

²Centre for Organismal Studies (COS), Heidelberg University, Im Neuenheimer Feld 360, 69120 Heidelberg, Germany

³Cluster of Excellence GreenRobust, Heidelberg University, Heidelberg, Germany

*Corresponding author: E-mail: pascal.hunziker@cos.uni-heidelberg.de.

†Authors contributed equally.

Accepted: April 30, 2026

Abstract

Tropaeolum tuberosum is a tuber-forming crop native to the Andes, valued for its nutritional content, pest resistance, and adaptation to high-altitude environments. Despite its contribution to food security and sustainable agriculture, it remains an underutilized species with scarce genomic resources, limiting genetic and functional studies. To address this gap, we generated a reference genome assembly for a European ex situ tetraploid accession of *T. tuberosum* using PacBio HiFi sequencing. The assembly spans 1.3 Gb in 1,805 contigs (contigs N50 = 32.2 Mb, longest contig = 60 Mb) and recovers 87% of the estimated genome size. We assessed assembly completeness and accuracy using Benchmarking Universal Single-Copy Orthologs, which detected 98.5% complete genes (21.7% single-copy, 76.8% duplicated), 0.7% fragmented, and 0.7% missing, demonstrating near-complete gene space recovery consistent with a high-quality tetraploid genome. To evaluate transferability, we resequenced a field-collected Colombian genotype of economic relevance using Oxford Nanopore technology. Read mapping showed 99.7% of primary alignments to the reference (weighted mean coverage = 16.4x, 96.1% at $\geq 5\times$), confirming broad sequence conservation between accessions and validating the suitability of the ex situ reference as an anchor for in situ germplasm. Repetitive elements accounted for 71.3% of the genome. Using ANNEVO, we annotated 56,354 high-confidence protein-coding genes, achieving 98.3% complete BUSCOs, a PSAURON score of 97.2, and 90.5% taxonomic consistency with the rosoid lineage (OMark). This high-quality reference genome establishes a foundational resource for comparative genomics, population genomics, and functional analyses on *T. tuberosum*, supporting future breeding and conservation of this important food resource.

Key words: PacBio HiFi, underutilized, orphan crop, cubio, mashua, isaño.

Introduction

Tropaeolum tuberosum Ruiz & Pav. ($2n = 4x = 52$), also known as *mashua*, *cubio*, or *isaño*, is a tuber-forming crop traditionally cultivated by smallholder farmers in local communities in the Andes mountains. It thrives under harsh conditions—including strong diurnal temperature shifts and nutrient-poor soils—producing nutrient-rich tubers with high levels of carbohydrates, proteins, vitamins, and diverse secondary metabolites

with antioxidant, anti-inflammatory, diuretic, and antimicrobial properties, including glucosinolates (Apaza Ticona et al. 2020; Coloma et al. 2022; Luera-Quñones et al. 2025). These traits make *T. tuberosum* a promising species for food security as well as resilient and sustainable agriculture.

T. tuberosum belongs to the genus *Tropaeolum* (family Tropaeolaceae, order Brassicales), a Neotropical lineage of roughly 90 herbaceous species that includes

Significance

T. tuberosum is an underutilized Andean tuber crop that combines highly nutritious tubers with remarkable disease resistance and UV tolerance, supporting Andean smallholder livelihoods under harsh high-altitude conditions. Beyond its role as a food crop, it is cultivated widely as an ornamental due to its abundant, brightly colored flowers closely resembling those of its relative nasturtium. Despite this dual relevance, genomic resources have been entirely lacking, limiting efforts to understand its domestication history, adaptation, and potential as a climate-resilient alternative to potato with higher nutraceutical properties. Our reference genome assembly provides the first genome-scale resource for *T. tuberosum*, creating a foundation for dissecting polyploid genome organization, key agronomic and nutritional traits, and interactions with biotic and abiotic stress. By enabling evolution, domestication, comparative genomics, breeding, conservation, and functional studies, this resource unlocks *T. tuberosum* as a model for resilient, diversified cropping systems globally.

ornamental nasturtiums and several tuber-forming taxa (Andersson and Andersson 2000). Its placement in the Brassicales is of particular biochemical interest: although glucosinolates are characteristic of this order, their diversity outside the Brassicaceae remains poorly characterized, highlighting the value of genomic data from nonbrassicaceous taxa (Agerbirk and Koch 2025; Bird et al. 2025).

Like other edible Andean root and tuber crops, *T. tuberosum* exhibits remarkable morphological and genetic diversity (Pissard et al. 2008; Fonseca Hernández and Márquez-Cardona 2024; Ruíz-Mateus et al. 2025), yet low commercial value and limited market integration are jeopardizing its conservation as a food resource. In contrast to globally important tuber crops such as potato, its improvement has relied primarily on traditional farmer selection. Thus, despite its agronomic and nutritional potential, *T. tuberosum* remains largely underutilized (Luera-Quiñones et al. 2025).

No reference genome has been available for *T. tuberosum*, and genomic resources remain scarce. Here we report a long-read-based, pseudo-haplotype-resolved genome assembly of a tetraploid accession of *T. tuberosum*. While a recent preprint described a genome for *Tropaeolum majus* (Friedhoff et al. 2024), this is the first assembly of *T. tuberosum*, providing a genomic framework for investigating specialized metabolism, tuber development, and environmental adaptation in a nonbrassicaceous Brassicales lineage, while laying the foundation for genomics-assisted improvement of this neglected crop.

Results and Discussion

We sequenced and assembled the genome of *T. tuberosum* Ruiz & Pav (Tropaeolaceae, Brassicales) accession BG HEID 007454 from the living collection of the Botanic Garden Heidelberg, Germany (Fig. 1a and b). Using single-molecule real-time sequencing technology from Pacific Biosciences, a total of 128.2 Gb of PacBio HiFi sequencing data were produced. *k*-Mer analysis

revealed an estimated monoploid genome size of 418.1 Mb, consistent with an autotetraploid genome architecture and an estimated heterozygosity rate of 4.7% (Fig. 1c). A genome survey based on *k*-mer frequency analysis of Illumina data estimated comparable metrics (Fig. S1). *k*-mer-based genome size estimates (418 to 497 Mb) were moderately smaller than the monoploid genome size inferred from flow cytometry ($2c = 2.48$ Gb) (Vesely et al. 2013), which assuming tetraploidy yields a monoploid size of 620 Mb. Such underestimation is commonly observed in repeat-rich polyploid genomes due to compression of repetitive and near-identical homeologous regions. *k*-Mer analysis of both HiFi and Illumina data supports an autotetraploid origin of *T. tuberosum*, with *aaab* forms (2.75%) exceeding *aabb* forms (1.72%) (Fig. 1c, Fig. S1), in line with the dominant cytological evidence (Johns and Neil Towers 1981).

Genome Assembly

Initial genome assembly using hifiasm yielded a draft primary assembly containing 2,189 contigs and a pseudo-haplotype-resolved assembly comprising two haplotypes (hap1, 1.21 Gb and hap2, 0.98 Gb) of unequal size, as expected for autotetraploid genome assembly (Tables S1 and S2). Because the four homeologous copies do not segregate into two discrete sequence clusters, phasing algorithms partition them based on local heterozygosity patterns. Bandage plots showed numerous small contigs with low depth (Fig. S2). Coverage distribution was bimodal, with a peak at 30 to 200 $\times$ representing the nuclear genome and a tail of high-coverage contigs (>200 $\times$, $n = 93$) (Fig. S3). In total, 292 contigs showed very low coverage (<10 $\times$). Kraken2 classification confirmed that all 292 low-coverage contigs originated from nontarget organisms (Table S3). Contaminants ($n = 292$), plastid-derived ($n = 60$), mitochondria-derived ($n = 30$), and repeat contigs ($n = 2$) were removed (384 contigs in total, accounting for 17.5% of the original assembly by count).

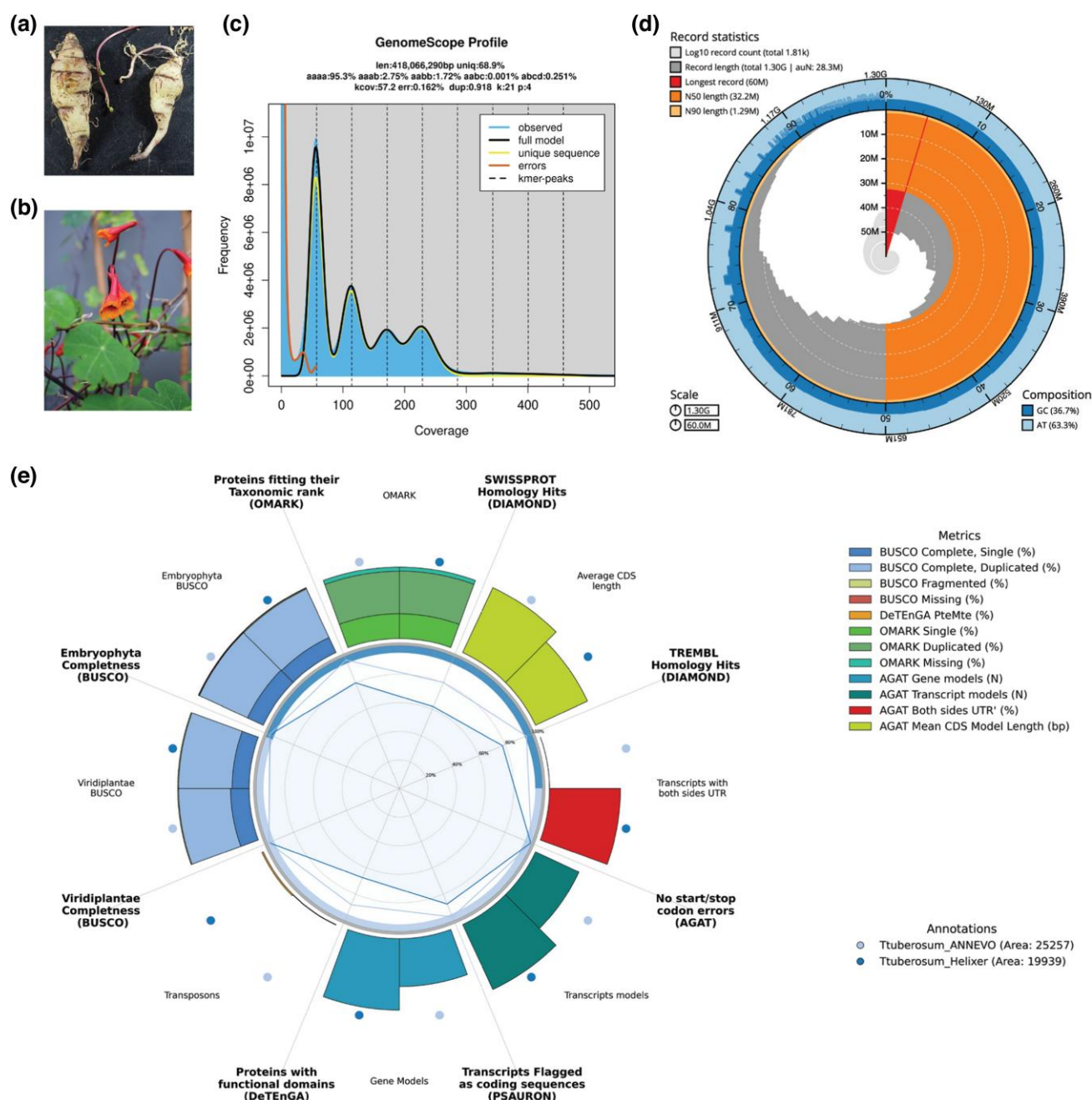


Fig. 1. The *Tropaeolum tuberosum* reference genome. a) Tubers of the BG HEID 007454 accession. b) Leaves, stems, and flowers of the BG HEID 007454 accession. c) Distribution profile of 21-mer analysis. d) Snail plot of the primary reference genome. e) Benchmarking Helixer and ANNEVO annotations.

The final primary assembly comprised 1,805 contigs (N50 of 32.2 Mb; largest = 60 Mb), with 90% of the assembly contained in 84 contigs (Fig. 1d, Table 1, Table S1). The size of 1.3 Gb matches the expected size of two monoploid complements (~1.24 Gb [Vesely et al. 2013]), consistent with a collapsed pseudo-haploid assembly representing most sequence diversity across four homeologs. BUSCO analysis showed 98.5%

completeness, 76.8% duplication, and 0.7% fragmentation, with identical results before and after decontamination, confirming no gene loss (Table 1, Table S1). Merqury analysis yielded a consensus quality value (QV) of 60.4 and *k*-mer completeness of 75.4% (Table 1, Table S4). The read-only and assembly copy number spectra aligned at ~56× coverage, as expected for a collapsed tetraploid, indicating that reduced

Table 1 Statistics for the *Tropaeolum tuberosum* reference genome assembly and annotation

Primary assembly	
Number of contigs	1,805
Total assembly size	1.30 Gbp
Largest contig	60 Mbp
N50 contig length	32.2 Mbp
L50 contig count	16
N90 contig length	1.3 Mbp
L90 contig count	84
GC content	36.70%
BUSCO v5	C: 98.5% [S: 21.7%, D: 76.8%], F: 0.7%, M:
embryophyta_odb10	0.7%, n: 1,614, E: 1.9%
QV	60.4025
k-Mer completeness	75.4455
Repeats	
LINE retroelements	54,466 (36.8%)
LTR retroelements	559,198 (33.6%)
DNA transposons	49,786 (1.9%)
Unclassified	680,448 (18.7%)
Small RNA	31,591 (13.2%)
Annotation (ANNEVO)	
Gene models	56,354
Exons	271,743
BUSCO v5	C: 98.3% [S: 24.2%, D: 74.2%], F: 0.7%, M:
embryophyta_odb10	0.9%, n: 1,614
PSAURON score	97.2
OMArk consistency results	Cons: 90.45% [P: 7.27%; F: 2.32%], Inco: 6.95% [P: 1.54%, F: 0.34%], Cont: 0.00%, Unkn: 2.60%
OMArk completeness results	malvids-HOGs: 11,704; S: 34.71%, D: 59.25% [U: 59.02%, E: 0.23%], M: 6.03%
OMArk species composition	rosids: 97.40%
Proteins with TREMBL hits	96.15
Proteins with SWISSPROT hits	80.39

multiplicity reflects intended homeolog collapse rather than sequence loss (Fig. S4). Assembly spectra exceeding read-only lines at higher multiplicities confirm comprehensive representation of multicopy sequences (Fig. S5).

To assess the utility of the assembly as a reference for natural populations, Nanopore long reads (~24x) from the field-collected genotype B15, originating from a farm in the department of Boyacá, Colombia, were mapped to the final primary assembly. A mapping rate of 99.7% and ≥98% of the assembly covered at a minimum depth of 5x demonstrates that the assembly captures the genomic content of geographically distant, in situ material with high fidelity, supporting its use as a reference for population-level and ecological genomic studies (Table S5).

Of the 1,805 assembly contigs, 1,009 (55.9%) were covered at ≥10x mean depth, and a further 288 (16.0%) at 5 to 10x (Table S5). The remaining 508 contigs (28.2%) showed low or absent coverage (<5x), collectively spanning only 23.5 Mb—1.8% of the nuclear assembly by length. Repeat density in this fraction (998 masked bp/kb) was indistinguishable from that of well-covered contigs (952 bp/kb), against a genome-wide nuclear repeat content of 71.3%, indicating that the low-coverage fraction is not an enrichment of repetitive sequence but rather reflects the structural properties of short, repeat-dense contigs that are inherently difficult to anchor by read mapping regardless of genotype. Consistent with this, gene density on low-coverage contigs (164 models/Mb) was comparable to that of well-covered contigs (128 models/Mb), and only 3,446 of 56,354 annotated gene models (6.1%) resided in this fraction (Table S5). The cross-genotype validation, therefore, confirms that the assembly provides comprehensive coverage of the genic and single-copy complement of the *T. tuberosum* genome, establishing it as a transferable reference resource for the species.

Together, the assembly statistics, BUSCO gene completeness, Merqury QV scores, k-mer completeness, and cross-genotype validation demonstrate that the *T. tuberosum* genome assembly is highly accurate, complete, and suitable for downstream analyses. Our primary assembly provides an annotation-grade reference genome sequence that was used for all gene prediction and functional annotation analyses described below.

The final haplotype-resolved assembly comprised hap1 (2,022 contigs, N50 = 18.4 Mb) and hap2 (580 contigs, N50 = 26.3 Mb), totaling 1.19 Gb and 0.96 Gb, respectively (Table S2). Combined, the 2.19 Gb assembly represents ~3.5x the monoploid genome size (620 Mb), recovering ~88% of the expected tetraploid complement (~2.48 Gb); the missing ~12% likely corresponds to highly repetitive or heterochromatic regions. Gene completeness analysis showed slightly lower BUSCO scores per haplotype but 98.8% completeness when merged, indicating that unassembled sequence is largely nongenic (Table S2). Merqury analysis revealed ~70% k-mer recovery per haplotype and 95% combined, with consensus QV 60 to 65, reflecting high sequence accuracy. The k-mer spectrum displayed four coverage peaks (~56x, ~112x, ~168x, ~224x), matching the expected 1x to 4x pattern for an autotetraploid at ~56x per haploid copy (Fig. S6). Minimal read-only fractions confirmed near-complete representation of the read k-mer space, while balanced hap1- and hap2-specific fractions and a dominant shared component at higher multiplicities (Fig. S7) are characteristic of autopolyploidy. Distinct haplotype-specific peaks verified successful

phasing, and per-haplotype spectra showed that missing *k*-mers reflect sequence assigned to the alternate haplotype rather than gaps (Figs. S8 and S9). Together, these results indicate a high-quality pseudo-haplotype-resolved, partially phased assembly of an autotetraploid genome.

De Novo Repeat and Gene Annotation

Using RepeatModeler, we identified 1,706 repeat families, including 732 structurally defined LTR retrotransposons (Table 1, Table S6). Repetitive elements accounted for 71.3% of the genome, dominated by Gypsy and Copia LTR retrotransposons. This composition aligns with other eudicot genomes of similar size (Chanderbali et al. 2022), where LTR retrotransposons contribute substantially to genome expansion (Zhang et al. 2020). The final nonredundant repeat library was used for genome-wide masking and annotation.

Helixer and ANNEVO predicted 84,059 and 56,354 protein-coding genes, respectively (Fig. 1e, Table 1, Table S7). Both annotations showed near-complete gene space coverage, with BUSCO completeness scores of 98.1% and 98.3%, respectively. In both cases, the majority of complete BUSCOs were duplicated (~75%), consistent with the retention of duplicated genes following ancient whole genome duplication events common in flowering plants (Van de Peer et al. 2017). Despite comparable completeness, ANNEVO consistently produced higher-quality gene models across all metrics assessed. ANNEVO achieved a substantially higher ORF accuracy score (PSAURON: 97.2 vs. 87.5), greater taxonomic consistency with the rosoid lineage (OMark: 90.5% vs. 71.5%), and a higher proportion of proteins with significant homology hits in both TrEMBL (96.2% vs. 78.2%) and Swiss-Prot (80.4% vs. 62.0%). Helixer's higher gene count was accompanied by clear signs of inflation: 4,557 overlapping gene models (vs. 59 in ANNEVO), 504 models with missing stop codons, 4.2% of proteins shorter than 50 amino acids (vs. 0.1%), and nearly four times more transposable element-derived transcripts (DeTenGA PteMte: 1.02% vs. 0.27%). The higher proportion of unknown hits in Helixer's OMark analysis (19.95% vs. 2.60%) further suggests that a substantial fraction of its additional gene models lack biological support.

Based on these results, the ANNEVO annotation was selected as the primary gene set. The 56,354 high-confidence protein-coding genes predicted by ANNEVO are consistent with the tetraploid genome of *T. tuberosum*. The closely related diploid *T. majus* harbors 32,924 protein-coding genes (Friedhoff et al. 2024). Assuming a similar ancestral gene complement, whole genome duplication would be expected to

approximately double this number, with subsequent gene loss and fractionation yielding the observed count in *T. tuberosum*.

Assembly and Annotation of Organelle Genomes

The plastid genome was assembled as two complete circular paths of identical length (~150 kb), representing the two orientations of the large single-copy/inverted repeat/small single-copy (LSC/IR/SSC) quadripartite structure characteristic of land plant chloroplasts (Fig. S10). This result is consistent with the well-documented flip-flop recombination of the IR region and confirms assembly completeness.

Mitochondrial genome assembly recovered one primary scaffold of 550,896 bp and a complex assembly graph with over 100 repeat patterns (Fig. S11). Such structural complexity is typical of plant mitochondria, which undergo repeat-mediated recombination and exist as dynamic genomic populations. Thirty high-coverage HiFi contigs mapped to this scaffold, confirming their mitochondrial origin and warranting their exclusion from the nuclear assembly. The most extreme outlier (ptg000094l, 1,293x) showed 33 BLASTn hits to the mitochondrial genome, indicating a highly reiterated region with elevated read depth.

Conclusions

This first reference genome assembly for *T. tuberosum* achieves high completeness and sequence quality, providing a foundation for comparative genomics, functional analyses, and diversity studies. This resource enables investigation of polyploid genome organization, tuber development, and specialized metabolism in *T. tuberosum* and comparative studies across the Tropaeolaceae family.

Materials and Methods

DNA in vitro-grown *T. tuberosum* accession BG HEID 007454 (Botanic Garden Heidelberg, Germany) was extracted using a CTAB-based protocol (Vaillancourt and Buell 2019). The Colombian B15 morphotype was collected from small producers' farms in Boyacá, Colombia (Fonseca Hernández and Márquez-Cardona 2024), and maintained in a conservation plot at San Javier Farm, Cogua, Cundinamarca, Colombia (05°03'37.0"N, 073°56'41.6"W; 2,590 m). DNA of B15 was extracted from fresh leaf tissue as described (Kalendar et al. 2021).

BG HEID 007454 DNA was sequenced on an Illumina NextSeq2000 at the Deep Sequencing Core Facility, Heidelberg University. Long-read sequencing used the SMRTbell Express Template Prep Kit 2.0 on a PacBio Revio by Novogene (<https://www.novogene.com/>).

B15 DNA libraries (SQK-LSK114) were sequenced on a PromethION 24 (V14 chemistry, R10 flow cells) by SNP Saurus (<http://www.snpsaurus.com>). Genome size, heterozygosity, and duplication rate were estimated using jellyfish v2.2.10 and GenomeScope v2.0.

HiFi reads were assembled with hifiasm v0.25.0-r726 using default settings. The primary assembly underwent contaminant (Kraken2 v2.1.6 using the standard RefSeq-based database [compiled 2025-09-24]; $n = 292$) and organelle (BLASTn v2.17.0; $n = 89$) removal. Quality metrics (gfastats v1.3.11, BUSCO v5.8.3 with the embryophyte_odb10 database, merquy v1.3, BlobToolKit v4.5.0) and B15 ONT mapping (minimap2 v2.30; mosdepth v0.3.13) are detailed in the [Supplementary Data](#).

The assembly was soft-masked (RepeatModeler v2.0.7, RepeatMasker v4.2.2). Gene models were predicted ab initio using Helixer v0.3.2 and ANNEVO v2.2.1 independently. Annotation quality was benchmarked with GAQET2 v2.0 including BUSCO v5.8.3 using the embryophyta_odb10 and viridiplantae_odb10 databases, OMARK v0.3.1, PSAURON v1.0.4, DeTenGA/TEsorter v1.0, and DIAMOND v2.1.11 against Swiss-Prot and TrEMBL.

Plastid and mitochondrial genomes were assembled from Illumina reads using GetOrganelle v1.7.7.1. The plastid was annotated with GeSeq. Full details are in the [Supplementary Data](#).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank the Botanical Garden Heidelberg for plant material (BG HEID 007454), Sandra Schmöckel and David Jarvis for helpful discussions, and Aureliano Bombarely for early ANNEVO testing. We thank David Ibberson (Deep Sequencing Core Facility, Heidelberg University) for technical support and Marcus Koch for commenting on the manuscript. We acknowledge the SDS@hd data storage service supported by MWK and DFG (INST 35/1803-1 FUGG, INST 35/1804-1 LAGG) and computational resources provided by the High Performance and Cloud Computing Group, University of Tübingen, through bwHPC and DFG (Project 455787709, bwForCluster BinAC2).

Author Contributions

Study conception and design: P.H. Sample preparation and experimental work: D.R.M., I.S., M.P.M., and P.H. Data analysis: D.R.M., W.T., and P.H. Manuscript draft: P.H. Manuscript final revision and consent: D.R.M., I.S., M.P.M., T.G., W.T., and P.H.

Funding

This work was funded by the Klaus Tschira Stiftung through the Klaus Tschira Boost Fund awarded to PH, and Fontagro (IDB) under grant “Root to Food: Improvement of Yield in Potatoes and Other Andean Tubers” (ATN/RF-18120-RG) to MPMC and WT.

Data Availability

Raw reads for the BG HEID 007454 and B15 genotypes used in this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession numbers PRJEB108055 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB108055>) and PRJEB109811 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB109811>), respectively. Reference genome assembly (v1.2), haplotype sequences, and annotation files have been deposited at FigShare (<https://doi.org/10.6084/m9.figshare.31338463>).

Ethics and Compliance

The *T. tuberosum* accession BG HEID 007454 was obtained under a Material Transfer Agreement from the Botanic Garden Heidelberg, Germany. Maintained in European collections since at least 1995 (predating the Nagoya Protocol and EU Regulation 511/2014), it lacks documented collecting event or provenance. The collection and access to the Colombian genetic resource *T. tuberosum* B15 were covered under the Access to Genetic Resources and Derived Products Agreement No. 256, “Studies on the Genetic Diversity of Native Crops of Colombia,” and its Addendum No. 1, signed between the Ministry of Environment and Sustainable Development of Colombia and Pontificia Universidad Javeriana.

Literature Cited

- Agerbirk N, Koch MA. Inconclusive evidence for chain-elongated methionine-derived glucosinolates beyond the family Brassicaceae. *Biochem Syst Ecol.* 2025;122:105033. <https://doi.org/10.1016/j.bse.2025.105033>.
- Andersson L, Andersson S. A molecular phylogeny of *Tropaeolaceae* and its systematic implications. *Taxon.* 2000;49:721–736. <https://doi.org/10.2307/1223973>.
- Apaza Ticona LN, Tena Pérez V, Bermejo Benito P. Local/traditional uses, secondary metabolites and biological activities of Mashua (*Tropaeolum tuberosum* Ruiz & Pavón). *J Ethnopharmacol.* 2020;247:112152. <https://doi.org/10.1016/j.jep.2019.112152>.
- Bird KA, Ramos AA, Kliebenstein DJ. Phylogenetic and genomic mechanisms shaping glucosinolate innovation. *Curr Opin Plant Biol.* 2025;85:102705. <https://doi.org/10.1016/j.pbi.2025.102705>.

- Chanderbali AS, et al. Buxus and Tetracentron genomes help resolve eudicot genome history. *Nat Commun.* 2022;13:643. <https://doi.org/10.1038/s41467-022-28312-w>.
- Coloma A, et al. Characterization of nutritional and bioactive compound in three genotypes of Mashua (*Tropaeolum tuberosum* Ruiz and Pavón) from different agroecological areas in Puno. *Int J Food Sci.* 2022;2022:7550987. <https://doi.org/10.1155/2022/7550987>.
- Fonseca Hernández LR, Márquez-Cardona MdP. Morphological and molecular characterization of cubios (*Tropaeolum tuberosum* Ruiz & Pavón) collected in two municipalities in Boyaca - Colombia. *Univ Sci (Bogota).* 2024;29:82–95. <https://doi.org/10.11144/Javeriana.SC291.mamc>.
- Friedhoff R, Wittstock U, Pucker B. 2024. Genome sequence of the medicinal plant *Tropaeolum majus* provides insights into flavonoid biosynthesis [preprint]. *bioRxiv* 621629. <https://doi.org/10.1101/2024.11.01.621629>.
- Johns T, Neil Towers GH. Isothiocyanates and thioureas in enzyme hydrolysates of *Tropaeolum tuberosum*. *Phytochemistry.* 1981;20:2687–2689. [https://doi.org/10.1016/0031-9422\(81\)85268-5](https://doi.org/10.1016/0031-9422(81)85268-5).
- Kalendar R, Boronnikova S, Seppänen M. Isolation and purification of DNA from complicated biological samples. *Methods Mol Biol.* 2021;2222:57–67. https://doi.org/10.1007/978-1-0716-0997-2_3.
- Luera-Quiñones S, Correa MJ, Moreno-Rojo C, Salvador-Reyes R, Paucar-Menacho LM. Mashua (*Tropaeolum tuberosum* Ruiz & Pavón): nutritional composition, bioactive compounds, and functional potential as an Andean natural ingredient. *Foods.* 2025;14:4198. <https://doi.org/10.3390/foods14244198>.
- Pissard A, Arbizu C, Ghislain M, Bertin P. Influence of geographical provenance on the genetic structure and diversity of the vegetatively propagated Andean tuber crop, mashua (*Tropaeolum tuberosum*), highlighted by intersimple sequence repeat markers and multivariate analysis methods. *Int J Plant Sci.* 2008;169:1248–1260. <https://doi.org/10.1086/591979>.
- Ruiz-Mateus D, Márquez-Cardona MdP, Terán W. Cubio tuber (*Tropaeolum tuberosum*, Tropaeolaceae) morphological diversity in Colombia and Bolivia. *Univ Sci (Bogota).* 2025;30:180–193. <https://doi.org/10.11144/Javeriana.SC30.cmd>.
- Vaillancourt B, Buell CR. 2019. High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing [preprint]. *bioRxiv* 783159. <https://doi.org/10.1101/783159>.
- Van de Peer Y, Mizrahi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet.* 2017;18:411–424. <https://doi.org/10.1038/nrg.2017.26>.
- Veselý P, Bureš P, Šmarda P. Nutrient reserves may allow for genome size increase: evidence from comparison of geophytes and their sister non-geophytic relatives. *Ann Bot.* 2013;112:1193–1200. <https://doi.org/10.1093/aob/mct185>.
- Zhang S-J, Liu L, Yang R, Wang X. Genome size evolution mediated by gypsy retrotransposons in Brassicaceae. *Genomics Proteomics Bioinformatics.* 2020;18:321–332. <https://doi.org/10.1016/j.gpb.2018.07.009>.

Associate editor: Vincent Castric